EXHIBIT 2

Exhibit 2

U.S. Patent No. 11,019,372 – Cisco Systems, Inc.

Claims 1, 6, 11

Scale Video Coding LLC ("SVC") provides evidence of infringement of claims 1, 6, and 11 of U.S. Patent No. 11,019,372 (hereinafter "the '372 patent") by Cisco Systems, Inc. ("Cisco"). In support thereof, SVC provides the following claim charts.

"Accused Instrumentalities" as used herein refers to at least the following Cisco products and services that implement the scalable video coding features of the AV1 coding format ("AV1"): Cisco Unified Communications Manager ("CUCM"); Cisco Webex (including Cisco Webex Cloud, Webex Teams, Webex Meetings, Webex App, Webex Meetings Server, Webex Meeting Center, Webex Desk Pro, Webex Codec Pro, Webex Codec Plus, Webex WebRTC client, Webex Control Hub, Webex Video Mesh, Webex Cloud-Connected UC); and Cisco Meeting Server version 3.9 and higher (including at least the CMS 1000 and CMS 2000 servers, and the Cisco Meeting Server Web App); along with associated hardware and/or software, including but not limited to other Cisco servers and related computer systems operated by Cisco that work in conjunction with CUCM, Cisco Webex, and Cisco Meeting Server.

These claim charts demonstrate Cisco's infringement by comparing each element of the asserted claims to corresponding components, aspects, and/or features of the Accused Instrumentalities. These claim charts are not intended to constitute an expert report on infringement. These claim charts include information provided by way of example, and not by way of limitation.

The analysis set forth below is based only upon information from publicly available resources regarding the Accused Instrumentalities, as Cisco has not yet provided any non-public information. An analysis of Cisco's (or other third parties') technical documentation and/or software source code may assist in fully identify all infringing features and functionality. Accordingly, SVC reserves the right to supplement this infringement analysis once such information is made available to SVC. Furthermore, SVC reserves the right to revise this infringement analysis, as appropriate, upon issuance of a court order construing any terms recited in the asserted claims.

SVC provides this evidence of infringement and related analysis without the benefit of claim construction or expert reports or discovery. SVC reserves the right to supplement, amend or otherwise modify this analysis and/or evidence based on any such claim construction or expert reports or discovery.

Unless otherwise noted, SVC contends that Cisco directly infringes the '372 patent in violation of 35 U.S.C. § 271(a) by selling, offering to sell, making, and/or using, the Accused Instrumentalities. The following exemplary analysis demonstrates that infringement. Cisco makes,

uses, sells, imports, or offers for sale in the United States, or has made, used, sold, or offered for sale in the past, without authority, without authority products, equipment, or services that infringe claims 1, 6, and 11 of the ’372 patent, including without limitation, the Accused Instrumentalities.

Unless otherwise noted, SVC believes and contends that each element of each claim asserted herein is literally met through Cisco’s provision of the Accused Instrumentalities. However, to the extent that Cisco attempts to allege that any asserted claim element is not literally met, SVC believes and contends that such elements are met under the doctrine of equivalents. More specifically, in its investigation and analysis of the Accused Instrumentalities, SVC did not identify any substantial differences between the elements of the patent claims and the corresponding features of the Accused Instrumentalities, as set forth herein. In each instance, the identified feature of the Accused Instrumentalities performs at least substantially the same function in substantially the same way to achieve substantially the same result as the corresponding claim element.

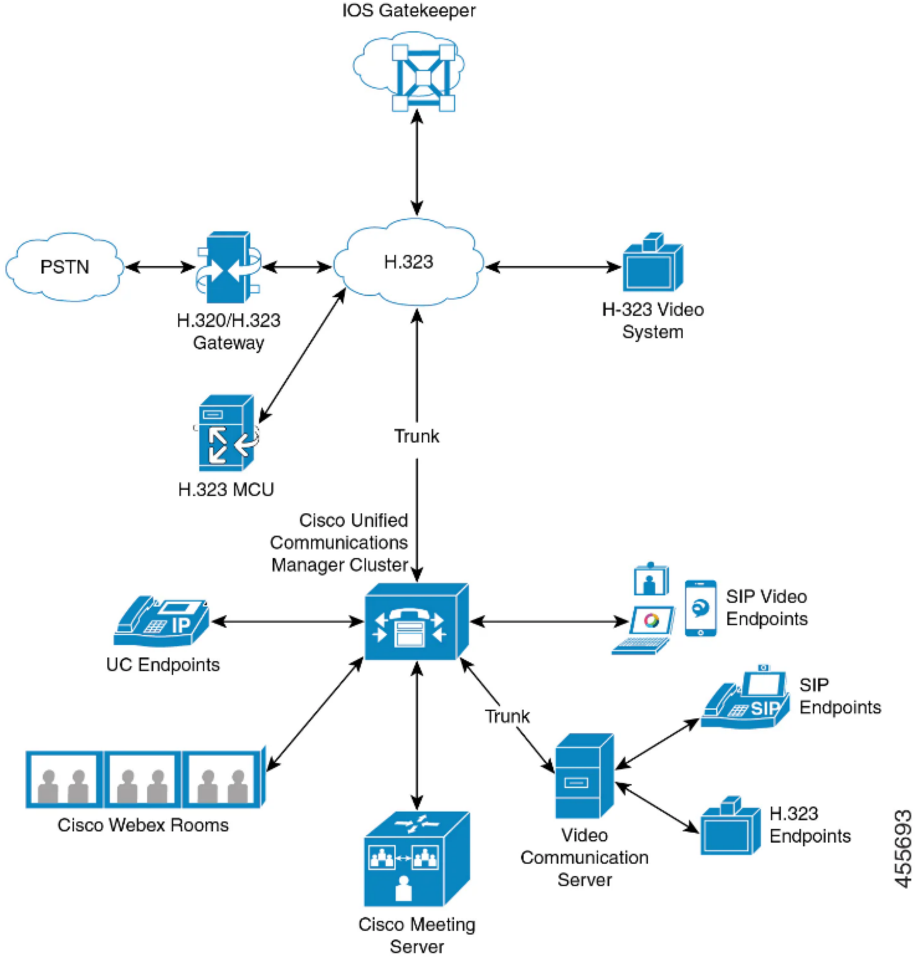
To the extent the chart of an asserted claim relies on evidence about certain specifically-identified Accused Instrumentalities, SVC asserts that, on information and belief, any similarly-functioning instrumentalities also infringe the charted claim. SVC reserves the right to amend this infringement analysis based on other products made, used, sold, imported, or offered for sale by Cisco. SVC also reserves the right to amend this infringement analysis by citing other claims of the ’372 patent, not listed in the claim chart, that are infringed by the Accused Instrumentalities. SVC further reserves the right to amend this infringement analysis by adding, subtracting, or otherwise modifying content in the “Accused Instrumentalities” column of each chart.

Claim Chart – Cisco

’372 Patent Claims 1, 6, 11	Accused Instrumentalities
1. A video router, comprising:	The Accused Instrumentalities include a video router. For example:

Video Network

The following illustration provides an example of a video network that uses a single Unified Communications Manager cluster. In a successful video network, any endpoint can call any other endpoint. Video availability only exists if both endpoints are video-enabled. Video capabilities extend across trunks.



	The Cisco video conference portfolio comprises the following video bridges: • Cisco TelePresence MCU series • Webex Meeting Server (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/14SU2/cucm_b_feature-configuration-guide-for-cisco14su2/cucm_m_video-telephony.html.) **Product Overview** Cisco Webex Meetings Server is a virtualized, software-based solution that runs on Cisco Unified Computing System™ (Cisco UCS®) servers and VMware. It uses virtual appliance technology for rapid turn-up of services to end users. With Cisco Webex Meetings Server, there are two options for enabling mobile users to more securely access Cisco Webex conferences without going through a VPN. The first option is to deploy reverse proxy (or edge servers) in the enterprise perimeter (or DMZ). The second option, shown in Figure 1, is to deploy the reverse proxy servers behind your internal firewall, thus eliminating all DMZ components and related information security concerns.

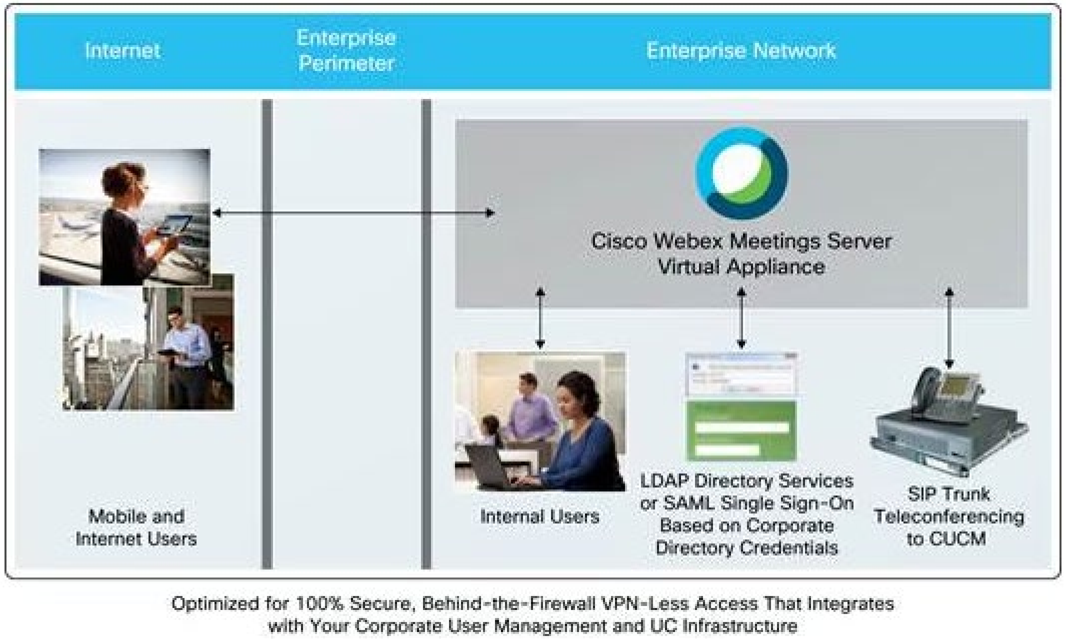


Figure 1.

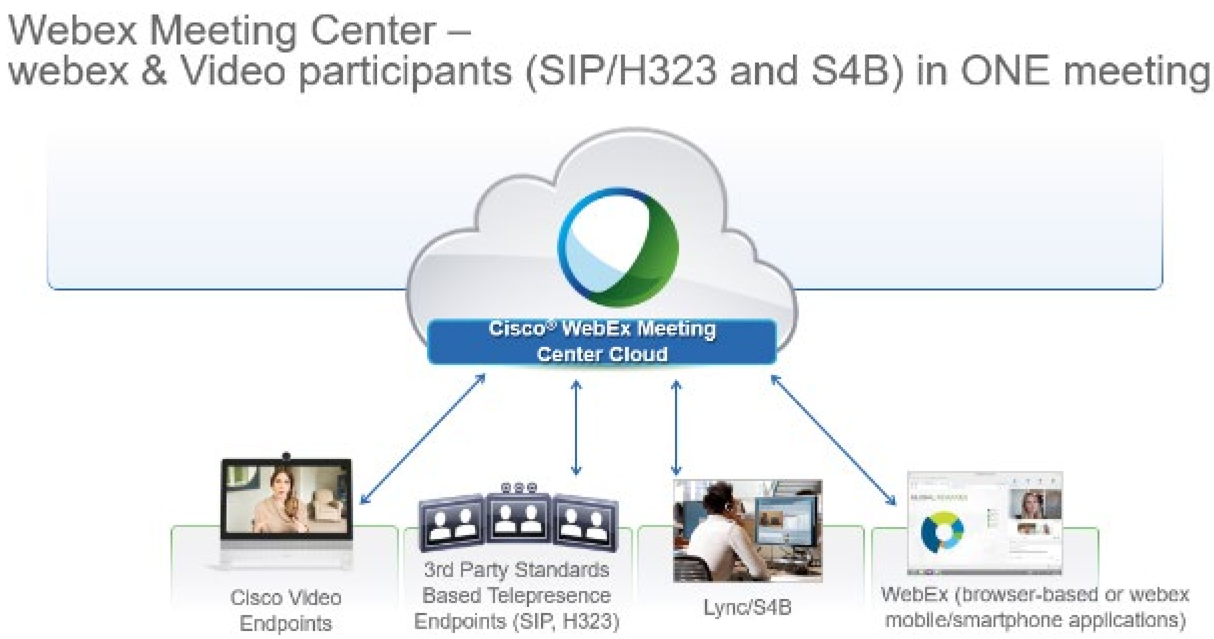
Full Deployment of Cisco Webex Meetings Server Behind a Firewall

System Requirements

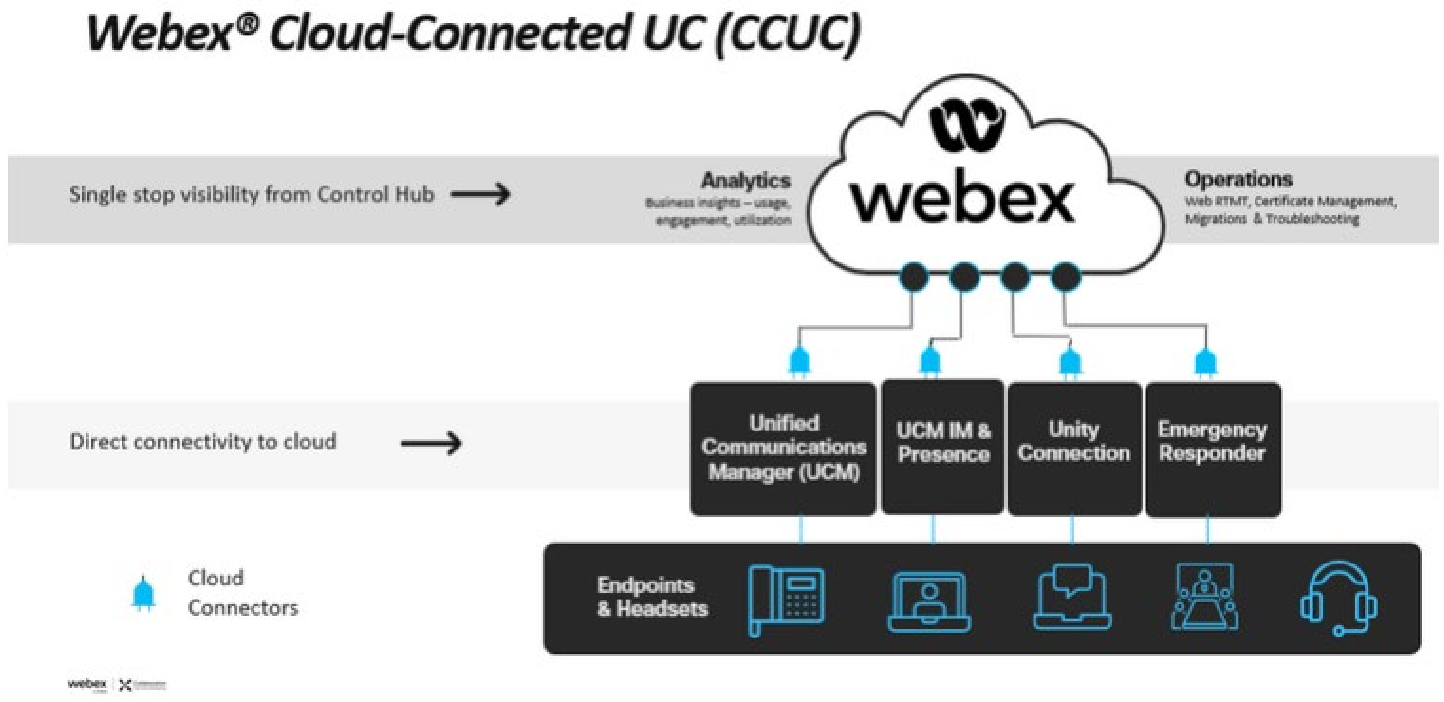
Cisco Webex Meetings Server is compatible with Cisco UCS servers that meet or exceed the specifications presented in this section.

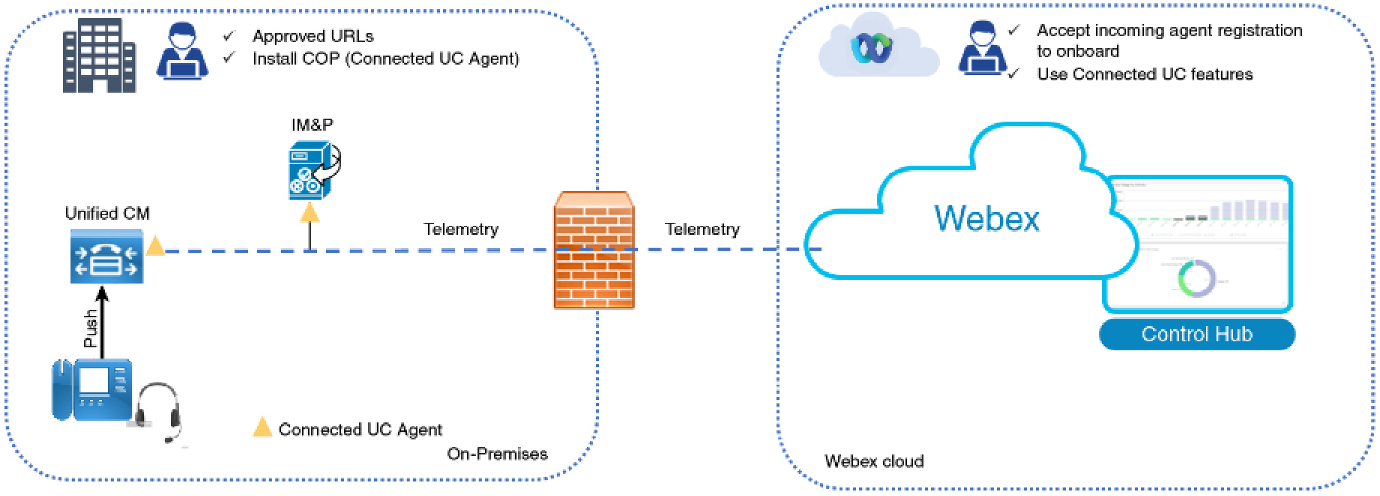
Module	Requirements
Host server	Cisco UCS C-Series rack server or equivalent B-Series blade server
Network interfaces	• Minimum 1 physical Network Interface Card (NIC) for a nonredundant configuration • Redundant configurations must have all NIC interfaces duplicated and connected to an independent switching fabric
Internal storage (direct attached storage [DAS]) for ESXi hosts where internal machines are deployed	Minimum of 4 drives in a RAID-10 or RAID-5 configuration
Internal storage (DAS)for ESXi hosts where Internet Reverse Proxy (IRP) virtual machines are deployed	Minimum of 2 drives in a RAID-1 configuration
SAN storage	Can be used as a substitute for DAS
Network-Attached Storage (NAS)	Can be used as a substitute for DAS or SAN
Hypervisor	• ESXi versions and vSphere licenses • 1 VMware license per processor socket
Email server	• Fully Qualified Domain Name (FQDN) of the mail server that the system uses to send emails • Port number: Default value of the Simple Mail Transfer Protocol (SMTP) port number is 25 or 465 (secure SMTP port number)

<table>
<tr>
<td></td>
<td>
(https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings-server/datasheet-c78-717754.html.)



(https://community.cisco.com/t5/collaboration-knowledge-base/cisco-webex-meetings-overview/ta-p/3648888.)
</td>
</tr>
</table>

	 (https://community.cisco.com/t5/collaboration-knowledge-base/webex-cloud-connected-uc-connect-your-cucm-to-webex-cloud-for/ta-p/4172103.)

<table>
<tr>
<td></td>
<td>


Cloud-Connected UC Deployment

(https://help.webex.com/en-us/article/jv0u1db/Webex-Cloud-Connected-UC-Overview.)
</td>
</tr>
</table>

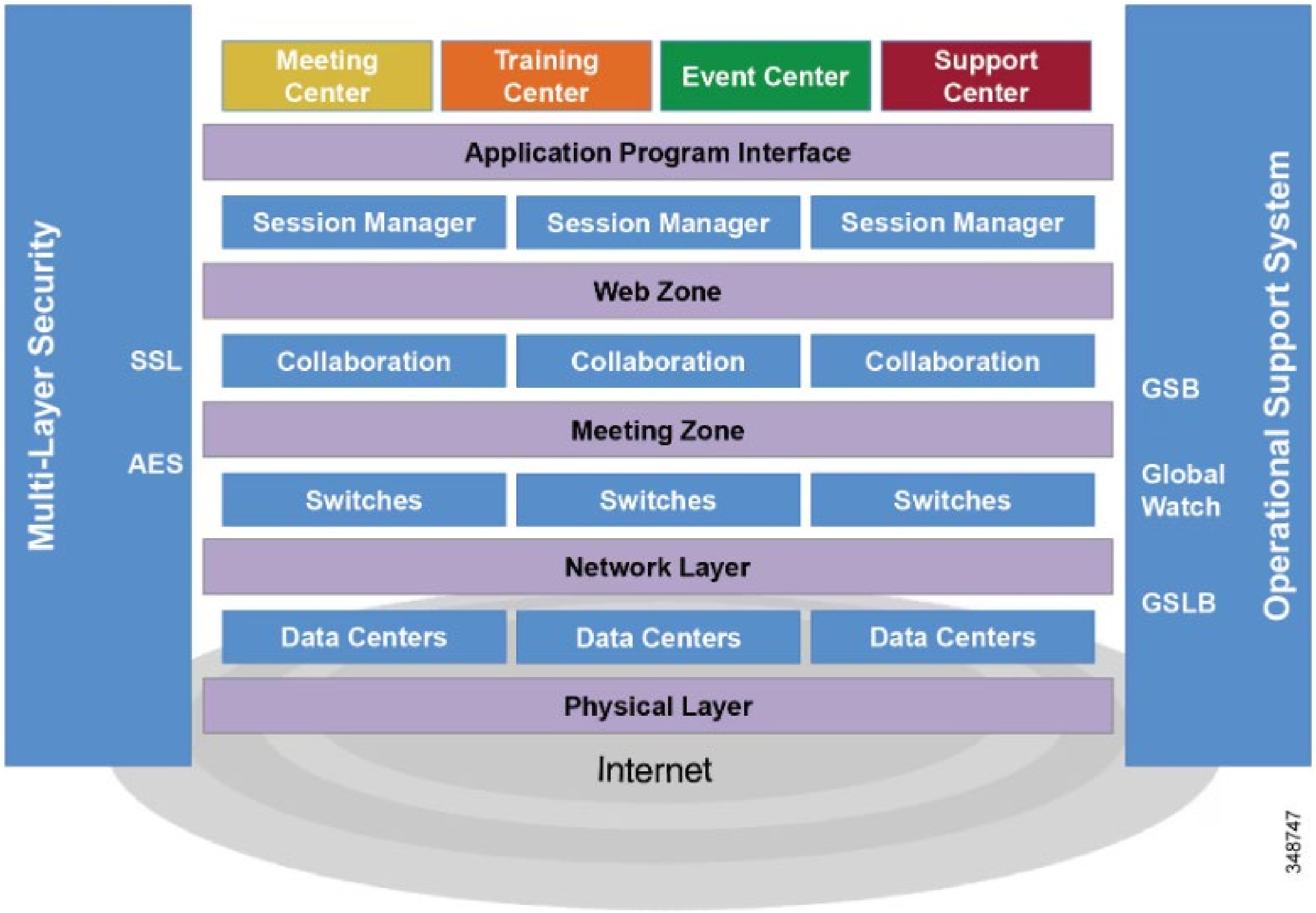
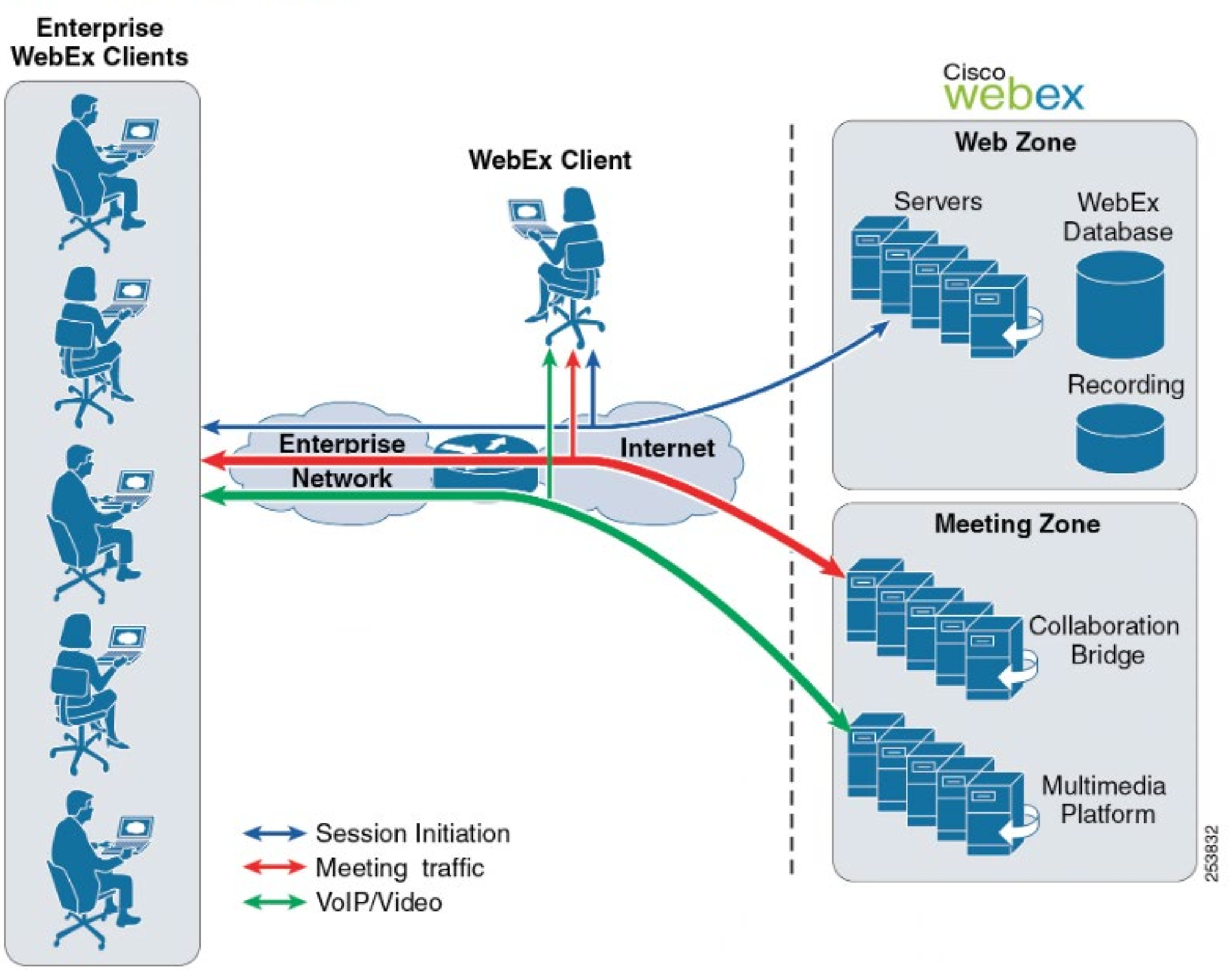
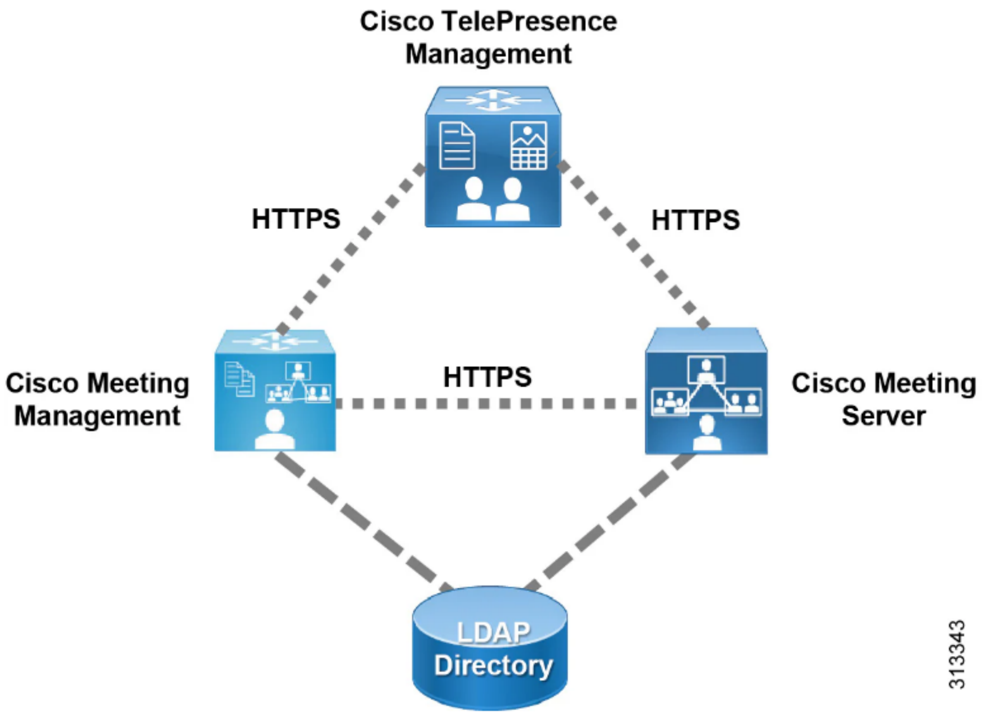
	Figure 11-14 Cisco WebEx Collaboration Cloud Architecture 

	Figure 11-15 WebEx Deployment 

	Cisco Rich Media Conferencing consists of the conferencing solutions described below. The details pertaining to each solution are described in each individual section that follows. • Cisco Unified CM Audio Conferencing This solution allows Unified CM to use its internal software component or external hardware digital signal processors (DSPs) as the resources to perform audio conferencing. • Cisco Meeting Server Cisco Meeting Server is an on-premises video conferencing solution. Each user has a personal Space that can be used to conduct meetings. Users can manage items such Space creation, adding members to a Space, and PIN creation from the Cisco Meeting App. • Cisco Collaboration Meeting Rooms Hybrid Cisco CMR Hybrid combines the on-premises video conference and the WebEx Meeting Center conference into a single meeting, which allows TelePresence and WebEx participants to join and share voice, video, and content. CMR Hybrid meetings can be either scheduled or non-scheduled. • Cisco WebEx Meeting Center Video Conferencing Cisco WebEx Meeting Center Video Conferencing (formerly Cisco Collaboration Meeting Rooms (CMR) Cloud) is an alternate conferencing deployment model that does not require any on-premises conferencing resources or management infrastructure. It supports both scheduled and non-scheduled meetings as well as TelePresence, audio, and WebEx participants in a single call, all hosted in the cloud. • Cisco WebEx Meetings Server Where cloud-based web and audio conferencing is not suitable, it is possible to use the on-premises WebEx Meetings Server solution. This product offers a standalone audio, video, and collaboration web conferencing platform. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab12/collab12/confernc.html.)

	Network Traffic Planning Network traffic planning for Cisco WebEx Meeting Center Video Conferencing consists of the following elements: • WebEx Clients bandwidth The WebEx meeting client uses Scalable Video Coding (SVC) technology to send and receive video. It uses multi-layer frames to send video, and the receiving client automatically selects the best possible resolution to receive video that typically requires 1.2 to 3 Mbps of available bandwidth. For more information regarding network traffic planning for WebEx clients, refer to the *Cisco WebEx Network Bandwidth* white paper, available at https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meeting-center/white_paper_c11-691351.html • Bandwidth for video device from enterprise to WebEx Cloud For optimal SIP audio and video quality, Cisco recommends setting up the video bandwidth for at least 1.5 Mbps per device screen in the region associated with the endpoint registering with Cisco Unified CM. For example, if a triple-screen device registers with Unified CM, video bandwidth of 4.5 Mbps should be allocated in the associated region. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab12/collab12/confernc.html.)

	Figure 3-3 Cisco Meeting Management Architecture 

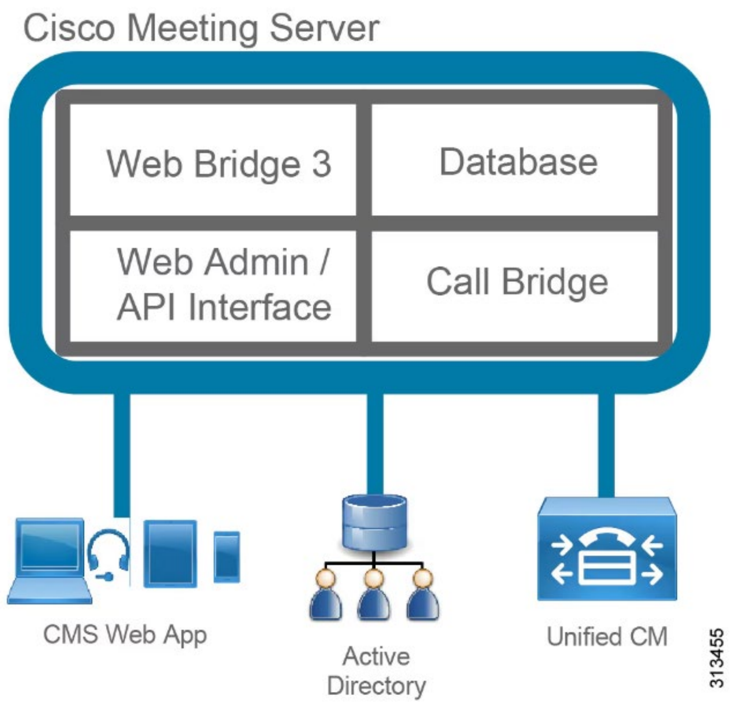
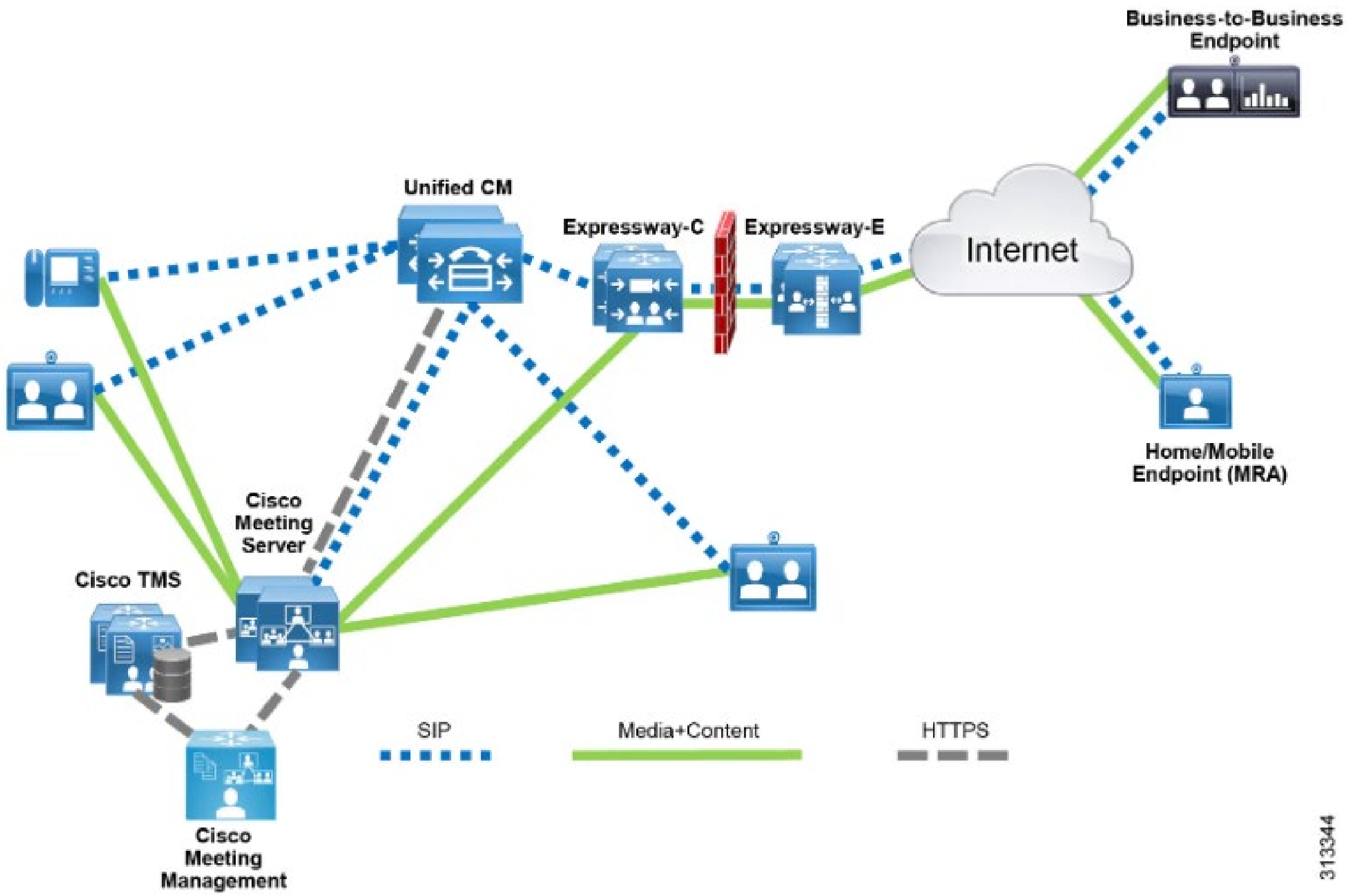
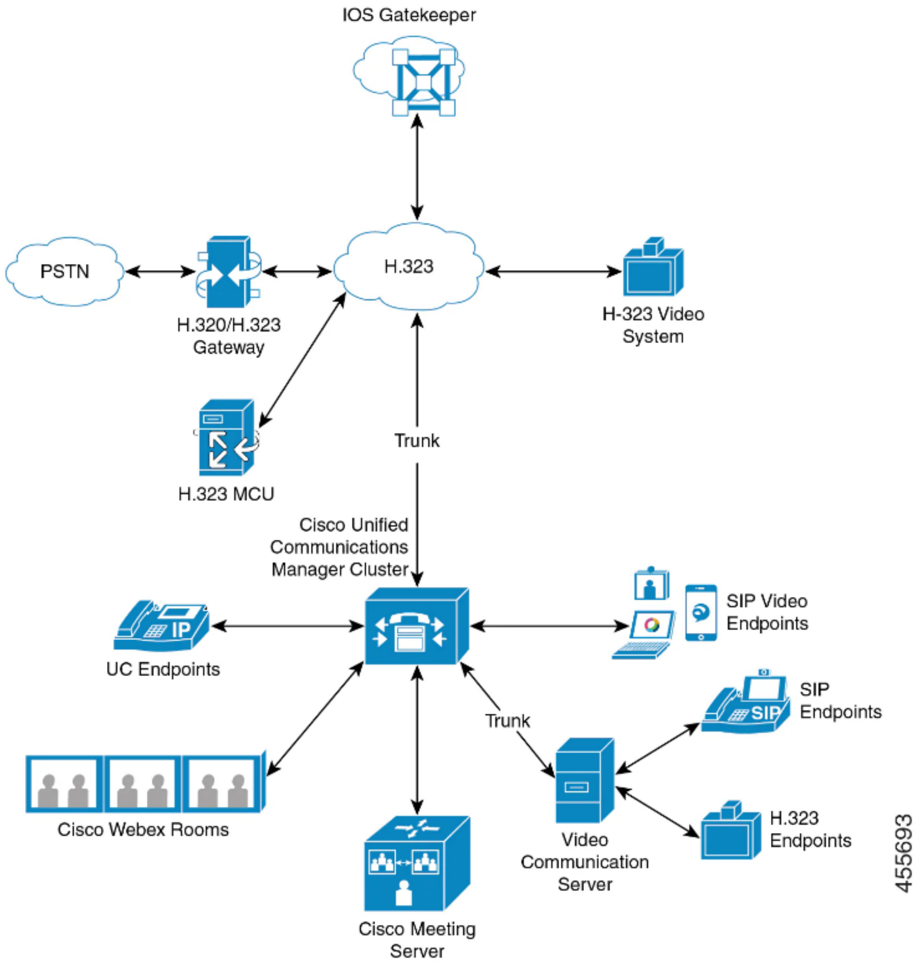
	 Cisco Meeting Server Web Bridge 3 Database Web Admin / API Interface Call Bridge CMS Web App Active Directory Unified CM 313455

	Figure 3-5 Standard Deployment  (https://www.cisco.com/c/en/us/td/docs/solutions/CVD/Collaboration/enterprise/14/collbcvd/conferencing.html.)

	How to deploy Cisco Meeting Server Talk with a Cisco salesperson or partner to learn about deployment and choose the best options for you. Select a platform Cisco Meeting Server software has been optimized to run on our UCS-based Cisco Meeting Server 1000 and Cisco Meeting Server 2000 Choose a licensing option Our multiparty option supports per-meeting licensing. Or you can purchase capacity units, as in a traditional license model. Consider add-on features Include recording ports, or consider Solution Plus partner Vbrick for recording/streaming distribution and Vyopta for assurance and analytics. Download the software Get the Cisco Meeting App for Macs and PCs on our site or from iTunes. Use the Apple Store for iOS devices. Download now › Go to iTunes › Platforms Cisco Meeting Server 1000 This Cisco UCS x86 server supports up to 120 simultaneous HD video conferencing calls. Cisco Meeting Server 2000 This Cisco UCS x86 server supports up to 875 simultaneous HD video conferencing calls. (https://www.cisco.com/c/en/us/products/conferencing/meeting-server/index.html.)
a memory; and a processor, wherein the processor executes instructions stored in the memory to cause the video router to:	The Accused Instrumentalities include a memory and a processor, wherein the processor executes instructions stored in the memory to cause the video router to [perform the claimed steps]. For example:

Video Network

The following illustration provides an example of a video network that uses a single Unified Communications Manager cluster. In a successful video network, any endpoint can call any other endpoint. Video availability only exists if both endpoints are video-enabled. Video capabilities extend across trunks.



	The Cisco video conference portfolio comprises the following video bridges: • Cisco TelePresence MCU series • Webex Meeting Server (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/14SU2/cucm_b_feature-configuration-guide-for-cisco14su2/cucm_m_video-telephony.html.)

Figure 11-14 Cisco WebEx Collaboration Cloud Architecture

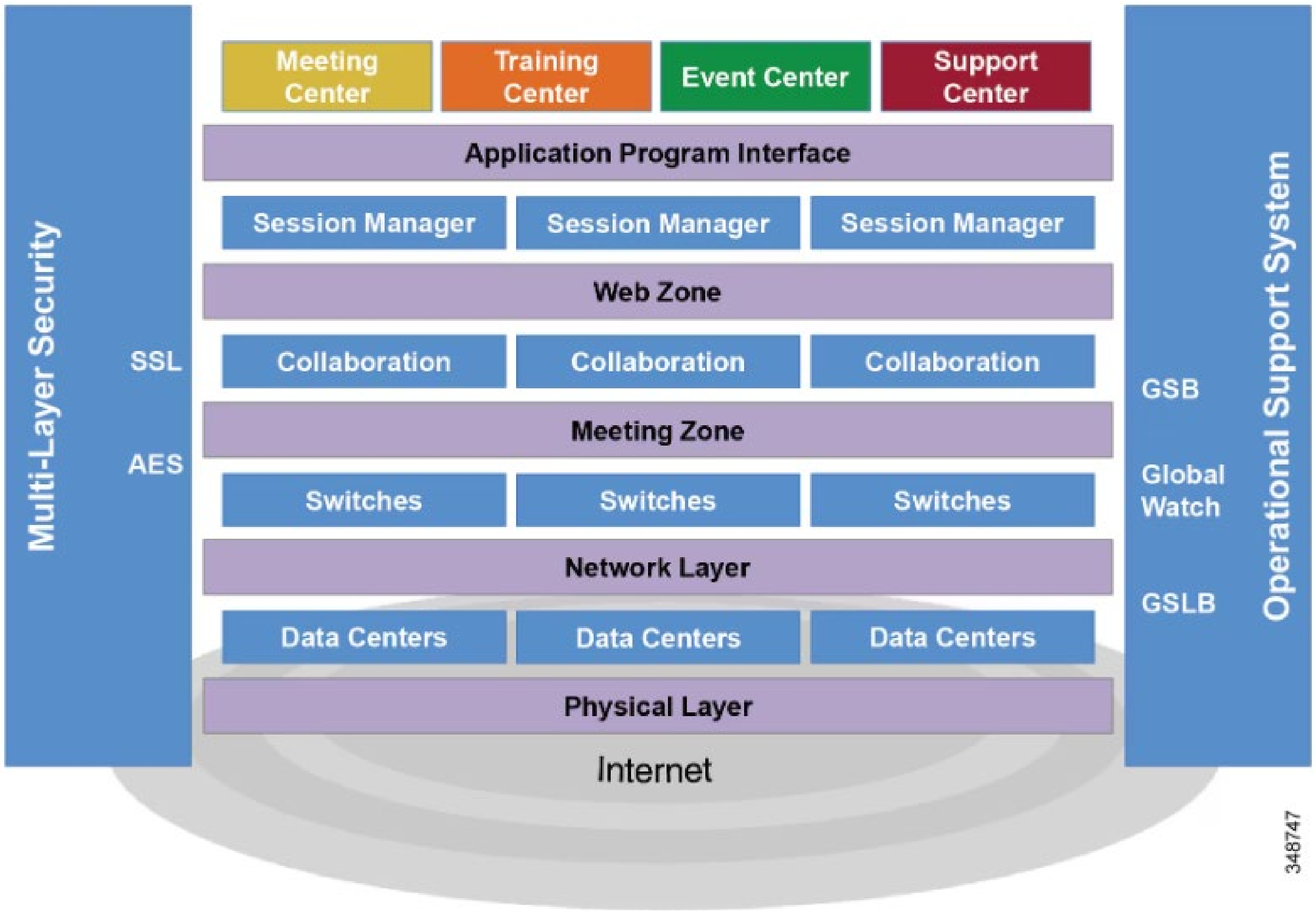
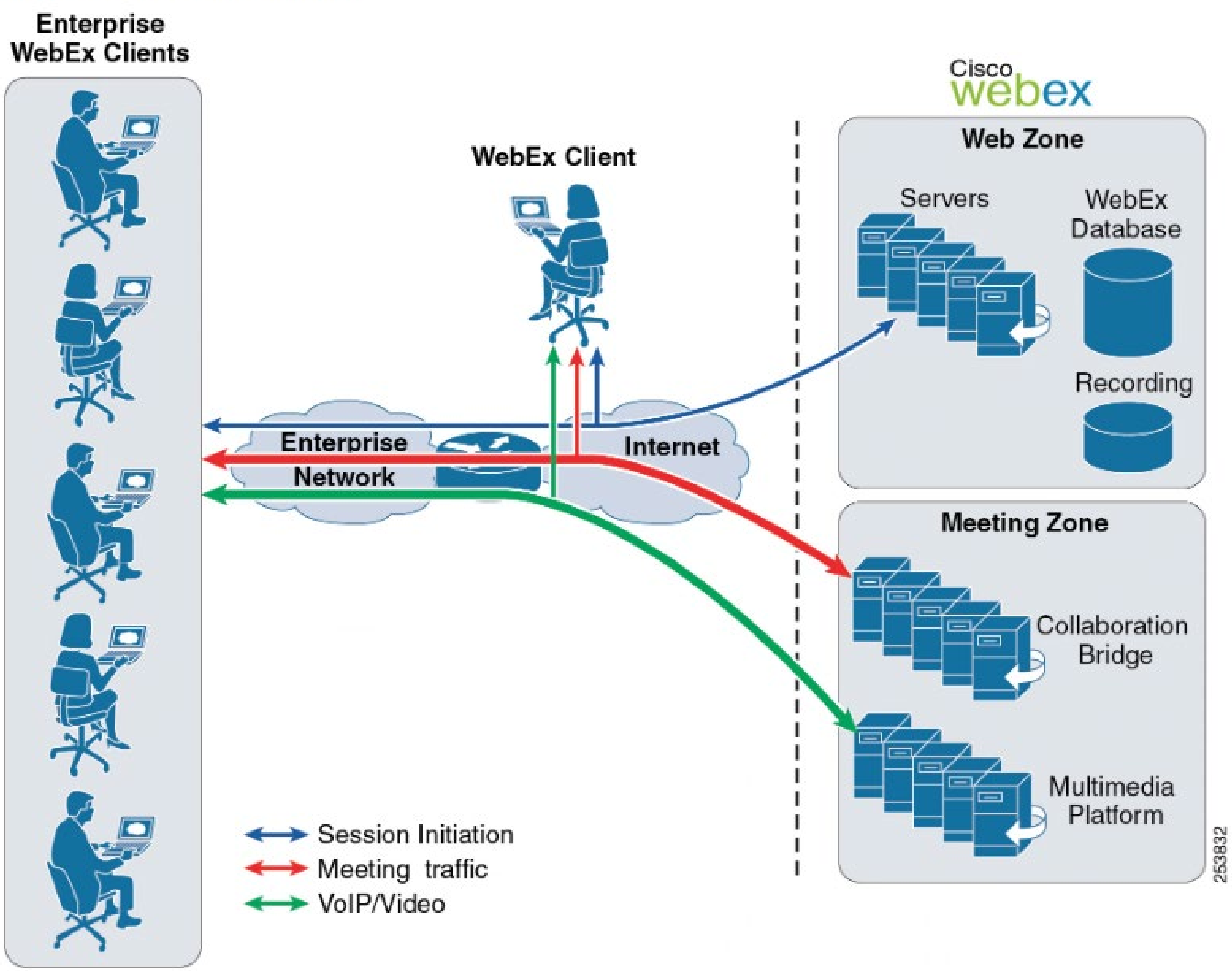


	Figure 11-15 WebEx Deployment  (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab12/collab12/confernc.html.)

Product Overview

Cisco Webex Meetings Server is a virtualized, software-based solution that runs on Cisco Unified Computing System™ (Cisco UCS®) servers and VMware. It uses virtual appliance technology for rapid turn-up of services to end users. With Cisco Webex Meetings Server, there are two options for enabling mobile users to more securely access Cisco Webex conferences without going through a VPN. The first option is to deploy reverse proxy (or edge servers) in the enterprise perimeter (or DMZ). The second option, shown in Figure 1, is to deploy the reverse proxy servers behind your internal firewall, thus eliminating all DMZ components and related information security concerns.

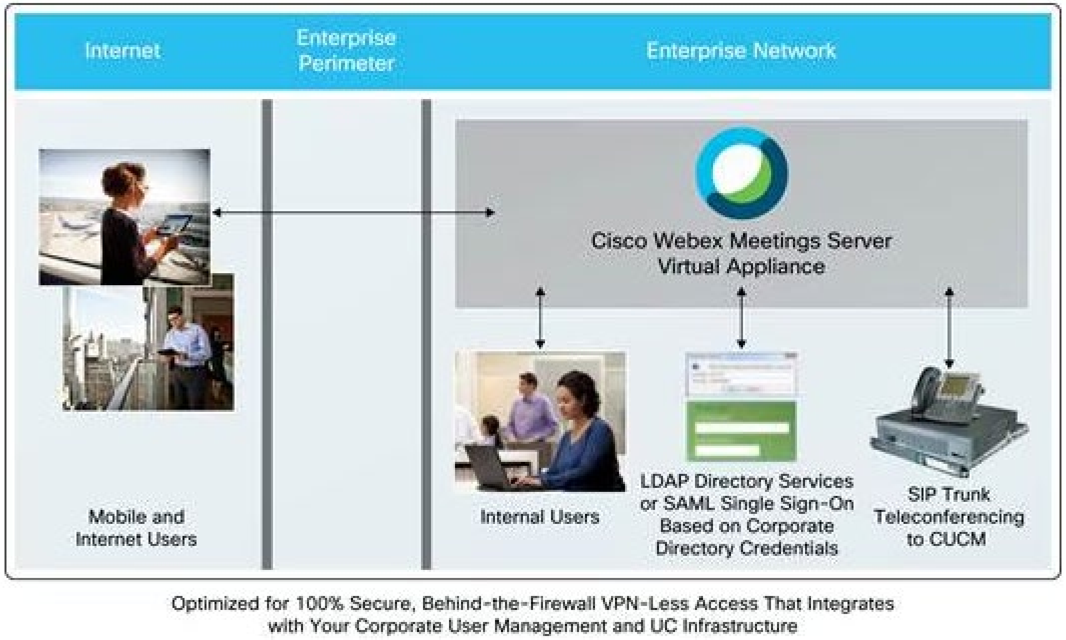


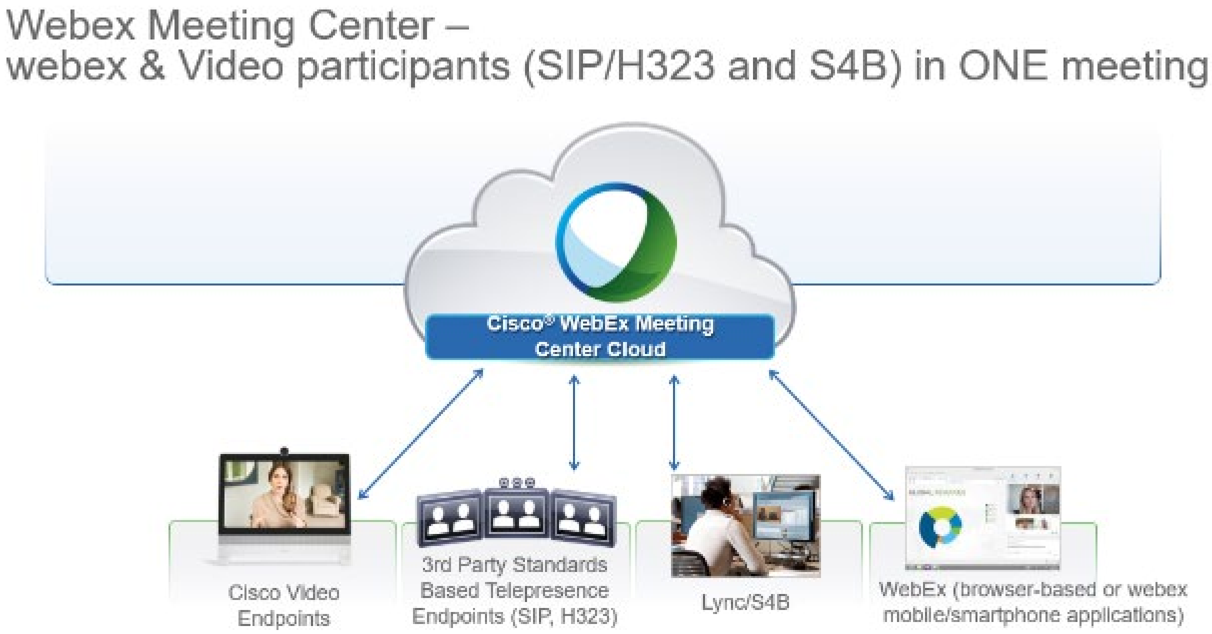
Figure 1.

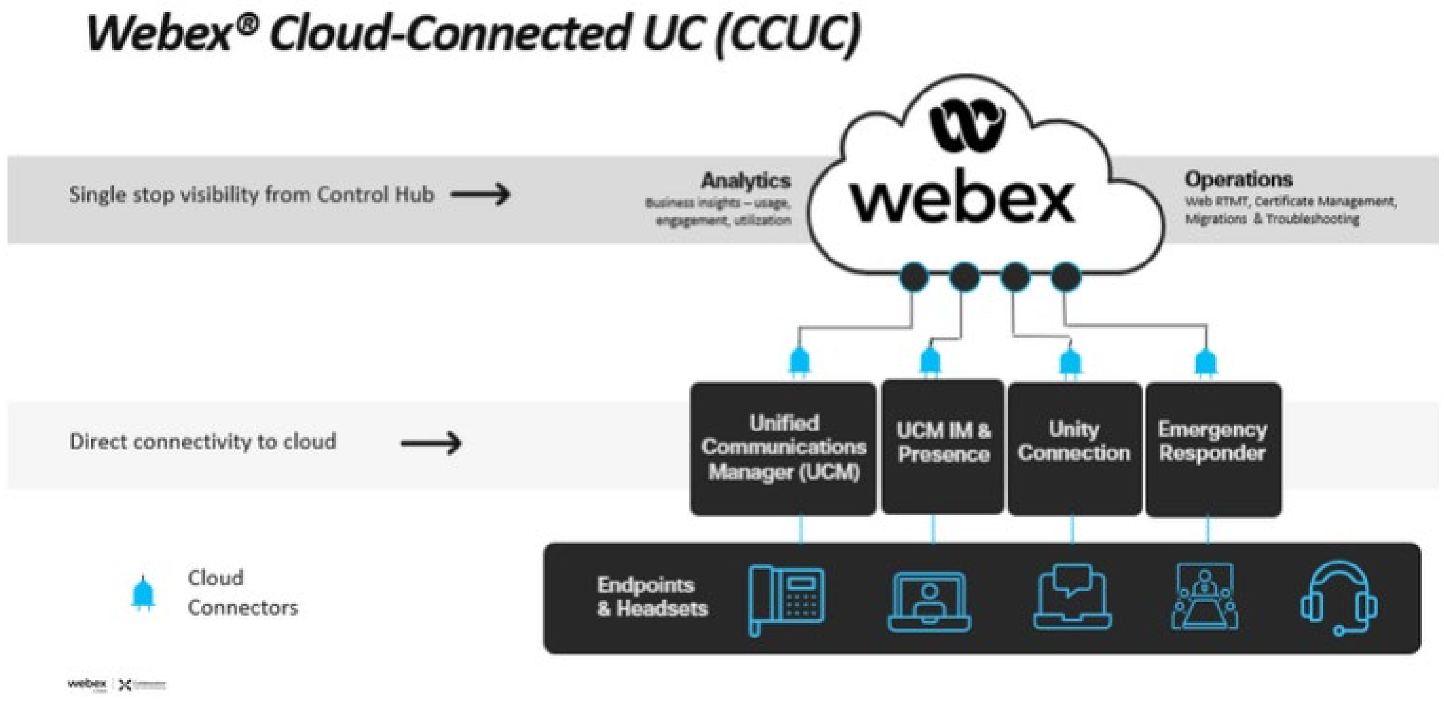
Full Deployment of Cisco Webex Meetings Server Behind a Firewall

System Requirements

Cisco Webex Meetings Server is compatible with Cisco UCS servers that meet or exceed the specifications presented in this section.

Module	Requirements
Host server	Cisco UCS C-Series rack server or equivalent B-Series blade server
Network interfaces	• Minimum 1 physical Network Interface Card (NIC) for a nonredundant configuration • Redundant configurations must have all NIC interfaces duplicated and connected to an independent switching fabric
Internal storage (direct attached storage [DAS]) for ESXi hosts where internal machines are deployed	Minimum of 4 drives in a RAID-10 or RAID-5 configuration
Internal storage (DAS)for ESXi hosts where Internet Reverse Proxy (IRP) virtual machines are deployed	Minimum of 2 drives in a RAID-1 configuration
SAN storage	Can be used as a substitute for DAS
Network-Attached Storage (NAS)	Can be used as a substitute for DAS or SAN
Hypervisor	• ESXi versions and vSphere licenses • 1 VMware license per processor socket
Email server	• Fully Qualified Domain Name (FQDN) of the mail server that the system uses to send emails • Port number: Default value of the Simple Mail Transfer Protocol (SMTP) port number is 25 or 465 (secure SMTP port number)

	(https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings-server/datasheet-c78-717754.html.)  (https://community.cisco.com/t5/collaboration-knowledge-base/cisco-webex-meetings-overview/ta-p/3648888.)

	 (https://community.cisco.com/t5/collaboration-knowledge-base/webex-cloud-connected-uc-connect-your-cucm-to-webex-cloud-for/ta-p/4172103.)

How to deploy Cisco Meeting Server

Talk with a Cisco salesperson or partner to learn about deployment and choose the best options for you.

Select a platform

Cisco Meeting Server software has been optimized to run on our UCS-based Cisco Meeting Server 1000 and Cisco Meeting Server 2000

Choose a licensing option

Our multiparty option supports per-meeting licensing. Or you can purchase capacity units, as in a traditional license model.

Consider add-on features

Include recording ports, or consider Solution Plus partner Vbrick for recording/streaming distribution and Vyopta for assurance and analytics.

Download the software

Get the Cisco Meeting App for Macs and PCs on our site or from iTunes. Use the Apple Store for iOS devices.

Download now › Go to iTunes ›

Platforms

Cisco Meeting Server 1000

This Cisco UCS x86 server supports up to 120 simultaneous HD video conferencing calls.

Cisco Meeting Server 2000

This Cisco UCS x86 server supports up to 875 simultaneous HD video conferencing calls.

(https://www.cisco.com/c/en/us/products/conferencing/meeting-server/index.html.)

Table 3. Ordering information

Platform (step 1)	Description
CTI-CMS-1K-M6-K9	Cisco Meeting Server 1000 M6
CTI-CMS-2K-M6-K9	Cisco Meeting Server 2000 M6
R-CMS-K9	Call Bridge Activation key for a third-party server or CMS 1K, only if not using Smart Licensing
R-CMS-2K-K9	Call Bridge Activation key for CMS 2K, only if not using Smart Licensing

(https://www.cisco.com/c/en/us/products/collateral/conferencing/meeting-server/datasheet-c78-742168.html.)

For example, the Accused Instrumentalities implement the AV1 standard. The AV1 standard discloses a method for transmitting video signals (e.g., video bitstream). The AV1 standard is a video coding & decoding standard. It discloses that the AV1 standard is developed for video bitstream communication application like Internet related application, streaming, etc. It induces transmission of encoded video bitstream via Internet.

	AV1 Bitstream & Decoding Process Specification *Last modified: 2019-01-08 11:48 PT* *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 (Title Page),* available at https://aomediacodec.github.io/av1-spec/av1-spec.pdf. Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.)

	 (http://dgql.org/~unlord/MHV2018.pdf.)

	## Video Calls The typical video call includes two or three Real-Time Protocol (RTP) streams in each direction (that is, four or six streams). The call can include the following stream types: • Video (H.261, H.263, H.263+, H.264-SVC, X-H.264UC, H.264-AVC, H.265, AV1 and VT Camera wideband video codecs) • Far-End Camera Control (FECC) - Optional • Binary Floor Control Protocols (BFCP) **AV1 Codec Support** AV1 is a next-generation video codec developed by the Alliance for Open Media. The benefits of AV1 are: • Reduced bandwidth consumption and better visual quality by utilizing better compression efficiency compared to other video encodings • Enables video for users on very low bandwidth networks • Significant screen sharing efficiency improvements over other codecs Unified Communication Manager supports negotiation of AV1 codec to establish media if endpoints support the AV1 codec. When both endpoints support Multiple Codecs in Answer, Unified CM negotiates all the matching codecs including AV1 based on the preference order received. The endpoint will then use one of the codecs from the negotiated codec list for media streaming. In a low bandwidth environment, the AV1 codec is preferred by the endpoint over other codecs in the negotiated list. When both the endpoints involved in the call do not support the Multiple Codec in Answer, and the AV1 is the preferred codec over other codecs, Unified CM selects AV1 as the negotiated codec.

	SIP Video SIP video supports the following video calls by using the SIP Signaling Interface (SSI): • SIP to SIP • SIP to H.323 • SIP to SCCP • SIP intercluster trunk • H.323 trunk • Combination of SIP and H.323 trunk SIP video calls also provide media control functions for video conferencing. Unified Communications Manager video supports SIP on both SIP trunks and lines support video signaling. SIP supports the H.261, H.263, H.263+, H.264 (AVC), H.264 (SVC), X-H.264UC (Lync), and AV1 video codecs (it does not support the wideband video codec that the VTA uses). (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/14SU2/cucm_b_feature-configuration-guide-for-cisco14su2/cucm_m_video-telephony.html.) **AV1 Codec Support** Unified Communications Manager now supports negotiation and passthrough of AVI codec. The AV1 is a modern codec that provides better compression and hence can provide the same user experience as H.264 video codec at half the bandwidth. AV1 codec will be supported by Cisco Webex Desk Pro Endpoint, Webex Codec Pro, and Room Panorama systems. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/rel_notes/14_0_1/cucm_b_release-notes-for-cucm-imp-14_0_1/cucm_m_new-and-changed-features.html.)

	webex ahead Collaboration Workspaces Customer Experience Event Management Innovation & AI Sear **The AV1 video codec comes to Webex!** On Dec 15, 2020 - By Webex Team | 4 Min Read Thomas Davies & Sijia Chen – Webex is rolling out the AV1 video codec into production early next year. This will bring our media quality to the next level. As a founding member of AOM, Cisco is proud to introduce this advanced video technology into the real time communications market It's here! We have begun the process of rolling out the advanced AV1 video codec across Webex, taking video quality to the next level in the process, and replacing the aging H.264 standard.

	What do I need to use AV1 in Webex? Transmitting AV1 is supported when sharing screens or applications with "Optimize for motion and video" selected , and when the machine you are on has at least four cores. Receiving AV1 is supported for any machine with at least two cores. AV1 will automatically be used for sharing this type of screen content whenever all participants in a meeting support it, otherwise it will automatically revert to H.264.

How we are rolling out AV1

Adopting a brand-new video codec has an impact on every part of our Collaboration portfolio, so we are going step-by-step.

In future releases we will systematically expand where we deploy AV1. The immediate next steps are to support AV1 for other desktop share modes – either optimized for text and images, or automatically optimized. AV1 works just as well for these modes too, but we are being careful to change things gradually to make sure the user experience is perfect at each step.

Webex employs a fundamentally switched architecture, where video from each participant in a meeting is coded on their machine at different qualities and sent via a server to the other meeting participants. Initially, if some of those participants cannot support AV1 then we will automatically fall back to using H.264. Over time we will also remove these restrictions by applying ad hoc transcoding between AV1 and H.264 for those participants. This will also allow AV1 meetings to be recorded without reverting to H.264, for example.

Mobile devices will also rapidly gain hardware AV1 support, and then AV1 can be rolled out to mobile too. Although our solution is software-based and very fast on ARM as well as x86 processors, it is always better to make use of hardware codecs where possible on mobile to get the best battery life possible.

We'll also be seeking to reduce the restrictions we have placed on core count for AV1 as we continue to optimize. In fact, remarkably, our AV1 solution uses little more CPU than H.264. However, there are a huge range of different machines out there, and again we are moving gradually to safeguard user experience.

(https://blog.webex.com/engineering/the-av1-video-codec-comes-to-webex.)

"16 years after H.264, it's time for something new. Today, we demo'd an industry first: live, real-time AV1 encoding and transmission in a Webex meeting, with HD video & screen share!" — Anurag Dhingra, Cisco Webex CTO

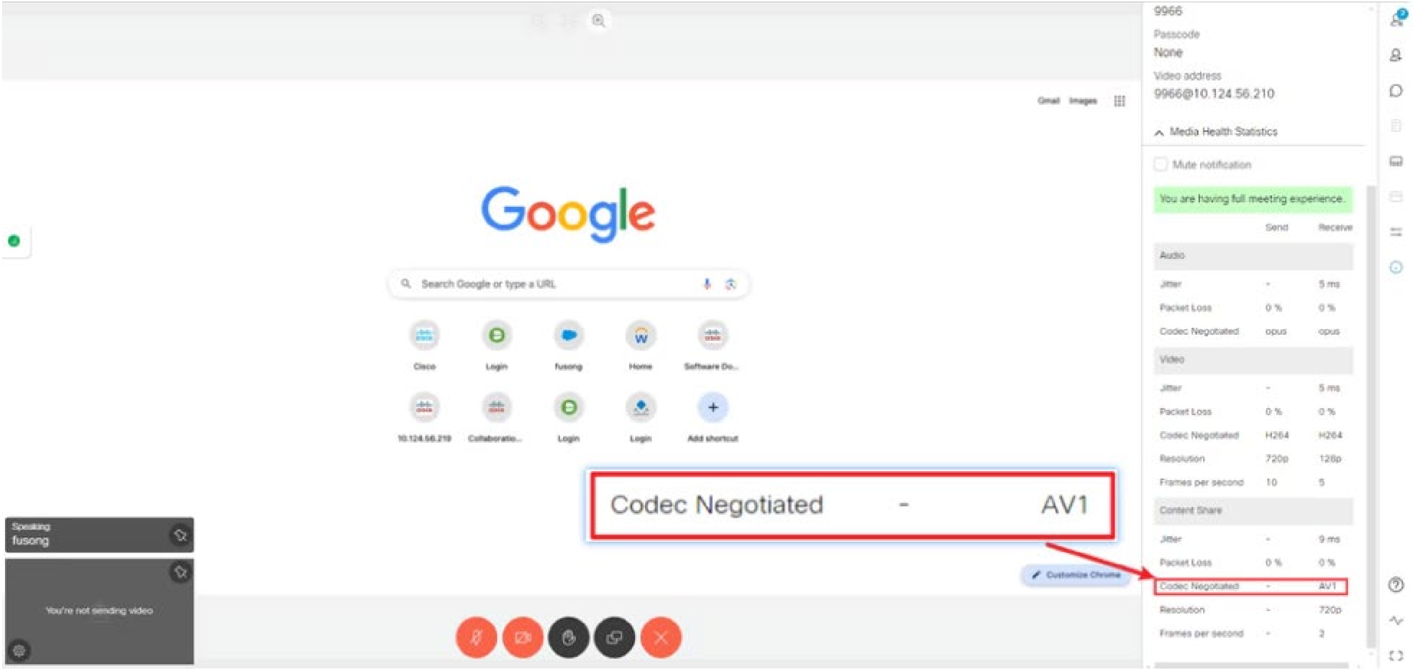
(https://medium.com/millicast/its-time-for-real-time-av1-video-encoding-withwebrtc-75a6aa64777c.)

AV1 Codec Support

Unified Communications Manager now supports negotiation and passthrough of AVI codec. The AV1 is a modern codec that provides better compression and hence can provide the same user experience as H.264 video codec at half the bandwidth. AV1 codec will be supported by Cisco Webex Desk Pro Endpoint, Webex Codec Pro, and Room Panorama systems.

See the compatibility matrix for the compatible version of Webex Room devices, Cisco Expressway, and Cisco Meeting Server.

(https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/rel_notes/14_0_1/cucm_b_release-notes-for-cucm-imp-14_0_1/cucm_m_new-and-changed-features.pdf.)

	2. The media health statistics of the content receiver show the content negotiate codec is AV1 on Chrome browser when receiving the content from CMS servers.  *Receiver content codec is AV1 on chrome* (https://www.cisco.com/c/en/us/support/docs/conferencing/meeting-server/221776-configure-av1-feature-on-cms.html.)
receive a layered video data stream including a base layer and a set of enhancement layers.	The Accused Instrumentalities receive a layered video data stream including a base layer and a set of enhancement layers. For example, the AV1 standard discloses receiving a layered video data stream (e.g., video bitstream) comprising a base layer (e.g., base layer) and a set of enhancement layers (e.g., enhancement layers).

	As shown below, the AV1 standard discloses an encoded video data bitstream using scalable video coding in a sequence of OBUs i.e., open bitstream unit. A metadata syntax of an OBU discloses scalability corresponding to the OBU. It discloses three types of scalabilities, Spatial scalability, Temporal scalability and Quality scalability. These scalabilities define a spatial layer having a corresponding spatial_id and a temporal layer having a corresponding temporal_id. Further, the AV1 standard discloses deriving a layered coded bitstream of base layer and enhancement layers using scalable video coding. It discloses a base layer having both spatial_id and temporal_id equal to zero and enhancement layers with either spatial_id or temporal_id equal to greater than zero values. **Bitstream** The sequence of bits generated by encoding a sequence of frames. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **5.8.1. General metadata OBU syntax**

metadata_obu( ) {	Type
metadata_type	leb128()
if (metadata_type == METADATA_TYPE_ITUT_T35)	
metadata_itut_t35( )	
else if (metadata_type == METADATA_TYPE_HDR_CLL)	
metadata_hdr_cll( )	
else if (metadata_type == METADATA_TYPE_HDR_MDCV)	
metadata_hdr_mdcv( )	
else if (metadata_type == METADATA_TYPE_SCALABILITY)	
metadata_scalability( )	
else if (metadata_type == METADATA_TYPE_TIMECODE)	
metadata_timecode( )	
}	

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 34 of 669)
	
	5.8.5. Metadata scalability syntax
	
	`metadata_scalability( ) {` — Type
	    `scalability_mode_idc` — `f(8)`
	    `if ( scalability_mode_idc == SCALABILITY_SS )`
	        `scalability_structure( )`
	`}`
	
	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

5.8.6. Scalability structure syntax

scalability_structure( ) {	Type
spatial_layers_cnt_minus_1	f(2)
spatial_layer_dimensions_present_flag	f(1)
spatial_layer_description_present_flag	f(1)
temporal_group_description_present_flag	f(1)
scalability_structure_reserved_3bits	f(3)
if (spatial_layer_dimensions_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1 ; i++) {	
spatial_layer_max_width[i]	f(16)
spatial_layer_max_height[i]	f(16)
}	
}	
if (spatial_layer_description_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1; i++)	
spatial_layer_ref_id[i]	f(8)
}	
if (temporal_group_description_present_flag) {	
temporal_group_size	f(8)
for (i = 0; i < temporal_group_size; i++) {	
temporal_group_temporal_id[i]	f(3)
temporal_group_temporal_switching_up_point_flag[i]	f(1)
temporal_group_spatial_switching_up_point_flag[i]	f(1)
temporal_group_ref_cnt[i]	f(3)
for (j = 0; j < temporal_group_ref_cnt[i]; j++) {	
temporal_group_ref_pic_diff[i][j]	f(8)
}	

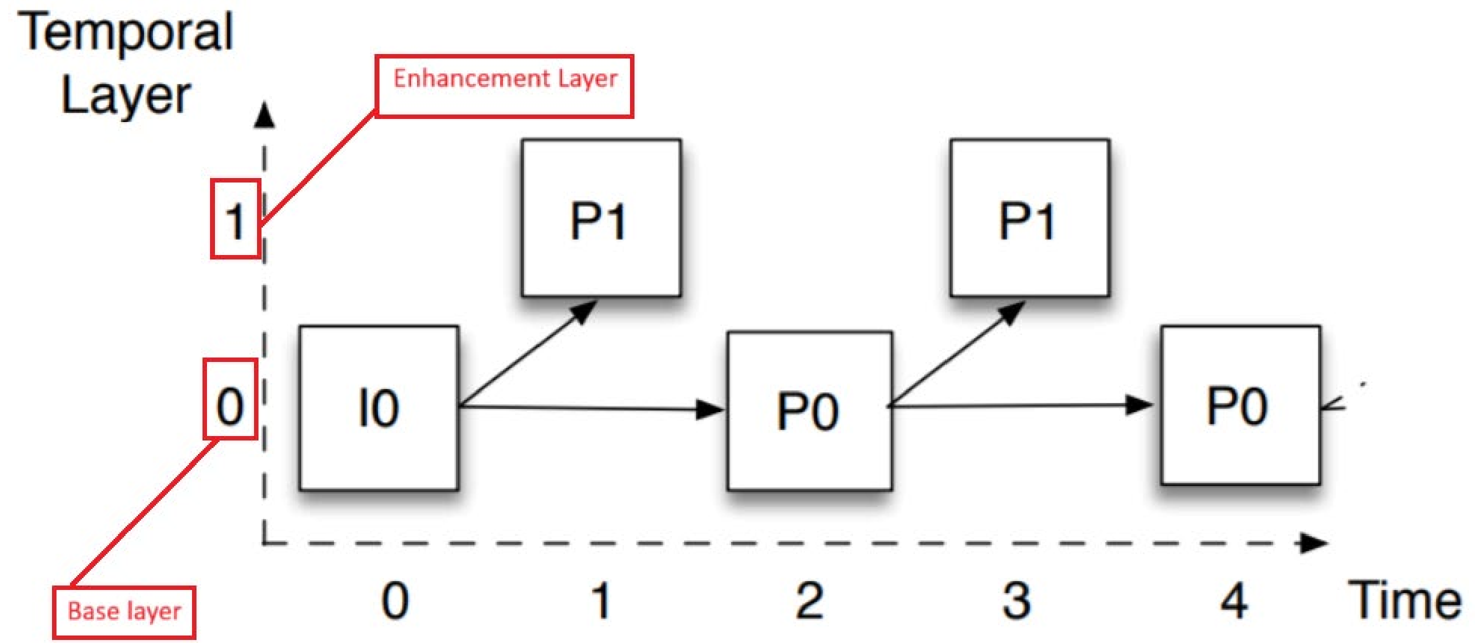
Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

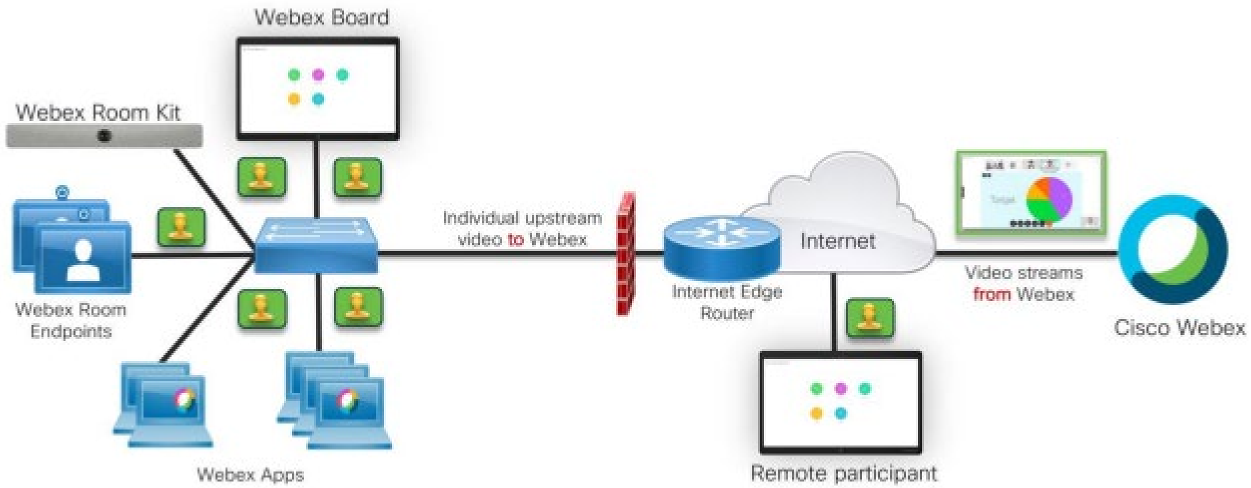
spatial_layer_ref_id[i] specifies the spatial_id value of the frame within the current temporal unit that the frame of layer i uses for reference. If no frame within the current temporal unit is used for reference the value must be equal to 255.

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669) Note that for a given picture, all frames follow the same inter-picture temporal dependency structure. However, the frame rate of each layer can be different from each other. The specified dependency structure in the scalability structure data must be for the highest frame rate layer. **temporal_group_temporal_id[i]** specifies the temporal_id value for the i-th picture in the temporal group. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669)* **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)*

	The AV1 codec can maintain up to eight reference frames, of which up to seven can be referenced by any new frame. AV1 also allows a frame to use another frame of a different spatial resolution as a reference frame. This allows internal resolution changes without requiring the use of key frames. These features together enable an AV1 encoder to implement various forms of coarse-grained scalability, including temporal, spatial, and quality scalability modes, as well as combinations of these, without the need for explicit scalable coding tools. Spatial and quality layers define different and possibly dependent representations of a single input frame. For a given spatial layer, temporal layers define different frame rates of video. Spatial layers allow a frame to be encoded at different spatial resolutions, whereas quality layers allow a frame to be encoded at the same spatial resolution but at different qualities (and thus with different amounts of coding error). AV1 supports quality layers as spatial layers without any resolution changes; hereinafter, the term "spatial layer" is used to represent both spatial and quality layers. This payload format specification provides for specific mechanisms through which such temporal and spatial scalability layers can be described and communicated. Temporal and spatial scalability layers are associated with non-negative integer IDs. The lowest layer of either type has an ID equal to 0. (https://aomediacodec.github.io/av1-rtp-spec/.) **Layer** A set of tile group OBUs with identical spatial_id and identical temporal_id values. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 3 of 669)* **Base layer** The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)*

	Enhancement layer A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)* As shown below, the AV1 standard discloses a base layer and enhancement layers for a coded video bitstream. **Note:** The term “spatial” refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* **Note:** Examples are given for non-scalable cases, but the constraints also apply to each operating point of a scalable stream. For example, consider a 60fps spatial scalable stream with a base layer at 960x540, and an enhancement layer at 1920x1080. The operating point containing just the base layer may be labelled as level 3.0, while the operating point containing both the base and enhancement layer may be labelled as level 4.1. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 641 of 669)*

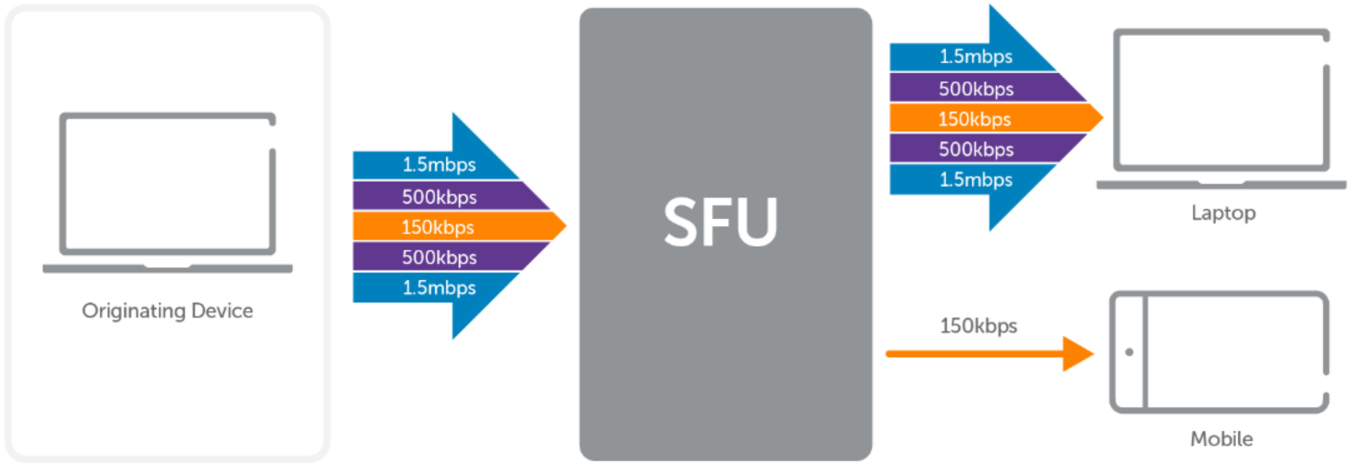
	 *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 126 of 669)*

	 ***Figure 11***: *Video Streams in a Webex Meeting* (https://www.cisco.com/c/dam/td-xml/en_us/collaboration/cloud_cmr/pcia_2_0/reports/Troubleshooting_Audio_and_Video_Quality_Using_Webex_Control_Hub.pdf.)
identify bandwidth-limited conditions of an internet protocol network between the video router and a plurality of video receivers,	The Accused Instrumentalities identify bandwidth-limited conditions of an internet protocol network between the video router and a plurality of video receivers. For example, the AV1 standard discloses identifying bandwidth-limited conditions (e.g., network conditions, available bandwidth condition for a receiving device, etc.) of an internet protocol network (e.g., Internet, etc.) between a video router (e.g., a video data transmitter such as a video bitstream encoder, etc.) and a plurality of video receivers (e.g., video data receivers such as a video bitstream decoder, etc.).

	Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) For example, one of the applications of the AV1 standard discloses a selective forwarding unit or middlebox (SFU or SFM) which analyses bandwidth condition between a publisher/transmitter (e.g., video router) and a playback

	device (e.g., video receiver). According to the bandwidth condition, it selectively forwards layers to the playback device. **How Does SVC Work?** Video Router 1. **Publisher**: Transmits a single media stream (with multiple video layers) to the Selective Forwarding Unit (SFU). 2. **SFU**: Examines the stream and selectively encodes it into high and low bitrate layers. It then distributes these layers based on each playback device's network limitations and available bandwidth. 3. **Playback Devices**: Receive the appropriate layers, ensuring optimal quality for each participant. Video Receiver (https://www.linkedin.com/pulse/scalable-video-coding-svc-sudeep-kumar-nofdc.)

	The SFU then sends the adjusted stream along to the appropriate **playback device**. Depending on the available bandwidth and other limitations of the target devices, some could receive a lower quality stream while others get the whole onion (so to speak).  (https://www.wowza.com/blog/scalable-video-coding-for-webrtc.) **Selective Forwarding Middlebox (SFM)** A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant.

	(https://aomediacodec.github.io/av1-rtp-spec/.) As another example, the Accused Instrumentalities are capable of identifying bandwidth-limited conditions, such as latency, jitter, and/or packet loss between backend servers and a plurality of user clients:

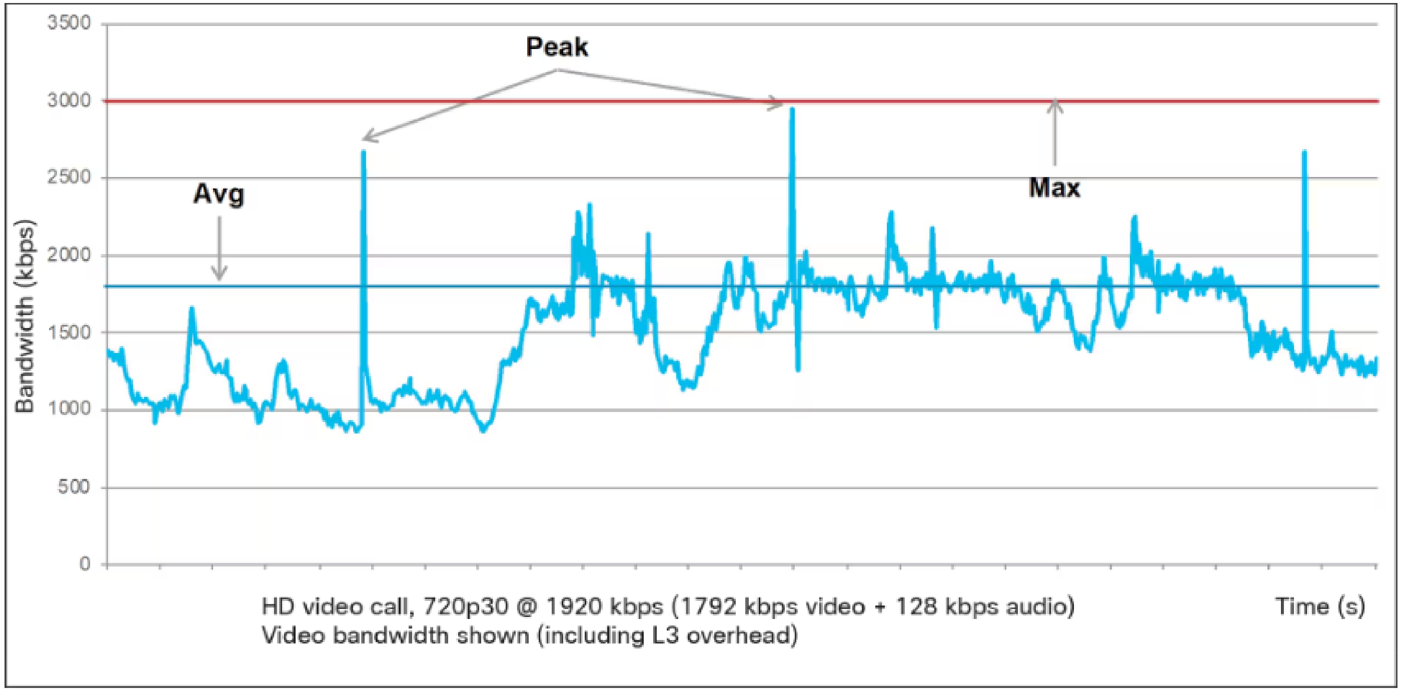
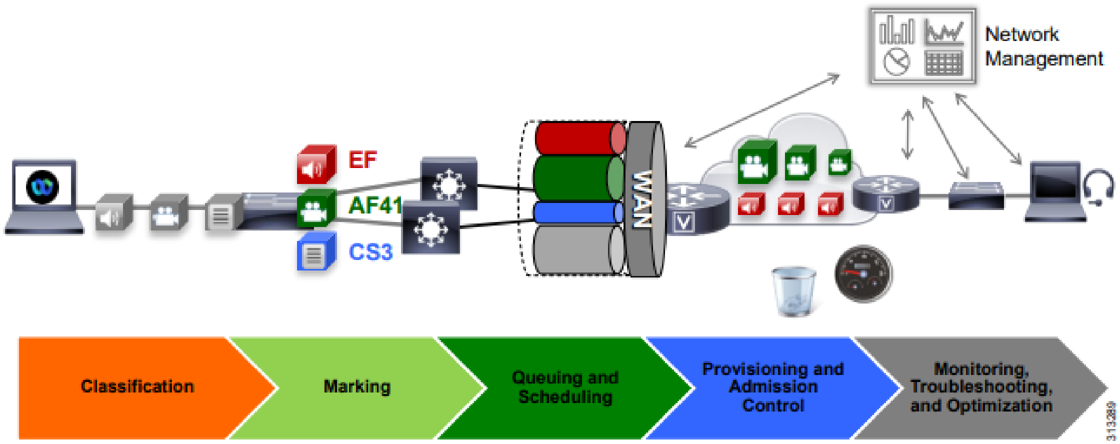
	Overview In this document we will discuss bandwidth utilization. Bandwidth values used will be in payload bit rate which does not include packetization overhead and are covered in 3 categories, average, peak and maximum bit rate: Average (**avg**) is the average over time for a meeting participant. Peak (**peak**) is the typical peak bursts over the same time period for a meeting participant. Maximum (**max**) is the maximum bit rate that the device is capable of either due to device limitations or device configuration.  **Figure 1.** Video Traffic: Bandwidth Usage High-definition Video Call (https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings/white_paper_c11-691351.html.)

Figure 1 Architecture for Bandwidth Management



Recommended Deployment

Modify the existing on-premises QoS switch and WAN and Internet policies to include Webex Services identification, classification, and marking.

- Identify and classify media and signaling traffic from the Webex App and associated workloads as well as Webex Devices.
- Media and signaling marking recommendations:
    1. Mark all audio with Expedited Forwarding class EF (includes all audio of voice-only calls as well as audio for all types of video calls).
    2. Mark all Webex App video with an Assured Forwarding class of AF41 for a prioritized video class of service and AF42 for an opportunistic class of service. The marking of AF41 or AF42 will depend on the choice of whether to deploy opportunistic video during the on-premises deployment phase.
    3. Mark all call signaling with CS3. (All call signaling in HTTPS traffic will be marked based on the enterprise's current policy of traffic marking for HTTP/HTTPS – unless using NBAR which is covered in detail below)
- Configure QoS on all media originating and terminating applications such as the Video Mesh Nodes, Expressway and Cisco Unified Border Element.
- Update the WAN edge ingress re-marking policy.
- Update the WAN edge egress queuing and scheduling policy.

Every Cisco video endpoint employs several smart media techniques to avoid network congestion, recover from packet loss, and optimize network resources. The following smart media techniques are just some of the techniques employed by Cisco Webex App and Devices:

Media resilience techniques

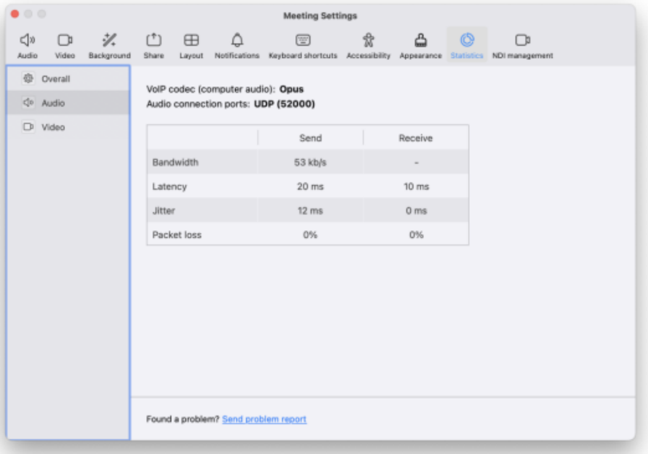
- Encoder pacing
- Forward Error Correction (FEC)
- Rate adaptation

(https://www.cisco.com/c/dam/en/us/td/docs/solutions/CVD/Collaboration/AltDesigns/BWM-Wbx.pdf.)

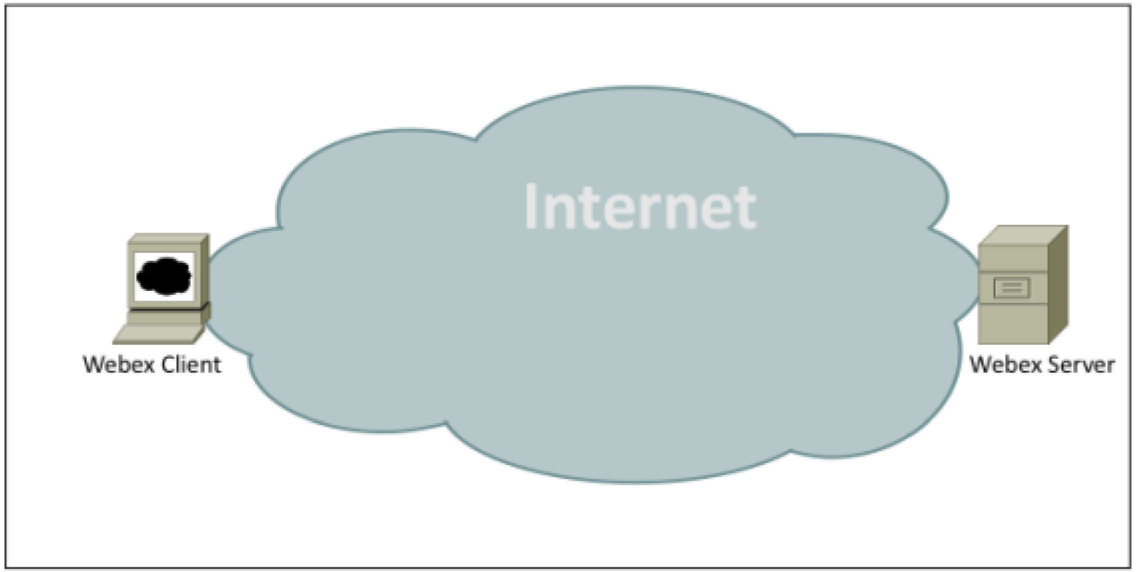
Poor Audio / Video Quality – Full-featured Meetings

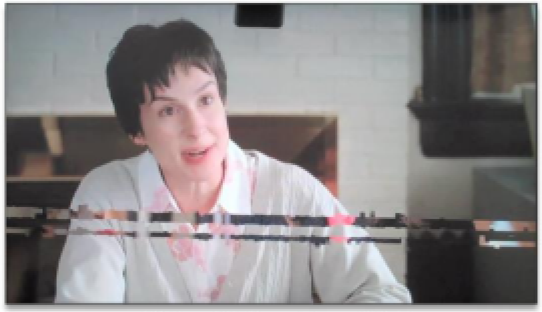
Help > Health Checker > Audio and Video Statistics...

- Indicates TCP or UDP w/ Source Port
- Latency / Packet Loss / Jitter



(https://www.ciscolive.com/c/dam/r/ciscolive/global-event/docs/2024/pdf/BRKCOL-3431.pdf.)

	There are three main factors that impact the quality of an audio or video call. These factors are **packet loss**, **latency**, and **jitter**. As shown in Figure 2, **packet loss is simply losing one or more packets within a stream of packets**. In this example, a Voice over IP (VoIP) packet is lost going from the Webex client to the Webex server. Loss can occur for a number of reasons in a network but most commonly it is caused by congestion or resource contention in the network or on the endpoints. This leads to routers, switches, or the endpoints themselves dropping or delaying packets. In other cases, packets encounter such a high delay that they are received too late to be played out. These packets will also be counted as lost.  ***Figure 2:*** *Packet Loss Example*

	Note: *Control Hub Diagnostics highlights potential problems with packet loss and delay by coloring voice and video streams as Good, Fair, or Poor. It is important to understand that just because Control Hub Diagnostics flags part of a conversation as Fair or Poor, it does not necessarily mean that the user had a bad experience. Voice and video compensation mechanisms and algorithms in combination with other factors can mitigate the impact of loss and delay to the point where it is not noticeable by the user.* **Video Quality Artifacts** The first step in troubleshooting video quality is to clearly understand the symptom experienced by the end user. Getting a screenshot or a picture of the display screen usually gives more information than a verbal description of the problem. The type of video artifacts observed in a screenshot or picture can be used to determine the troubleshooting path. For example, in Figure 13, pictures (a) and (b) are artifacts that are caused when video streams experience packet loss. Just by seeing the picture, you can immediately start troubleshooting to identify the source of the packet loss.  a. Video with line stripe artifacts  b. Frozen video with block artifacts

Webex Control Hub makes it quite easy to find the device type and operating system version used by the end user to join a Webex meeting. All you need is the end user's email address and meeting time to find the right meeting and to get end user's device details. Figure 14 shows the list of meetings attended by a participant with the email address, rtpmsuser1@gmail.com.

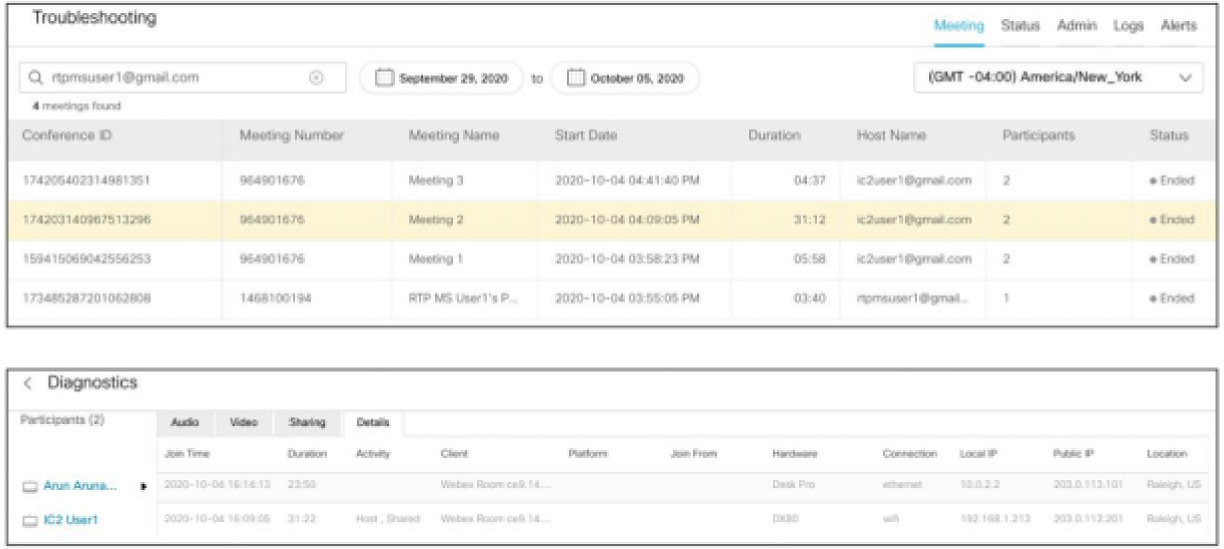


***Figure 14**: Participant Device Details*

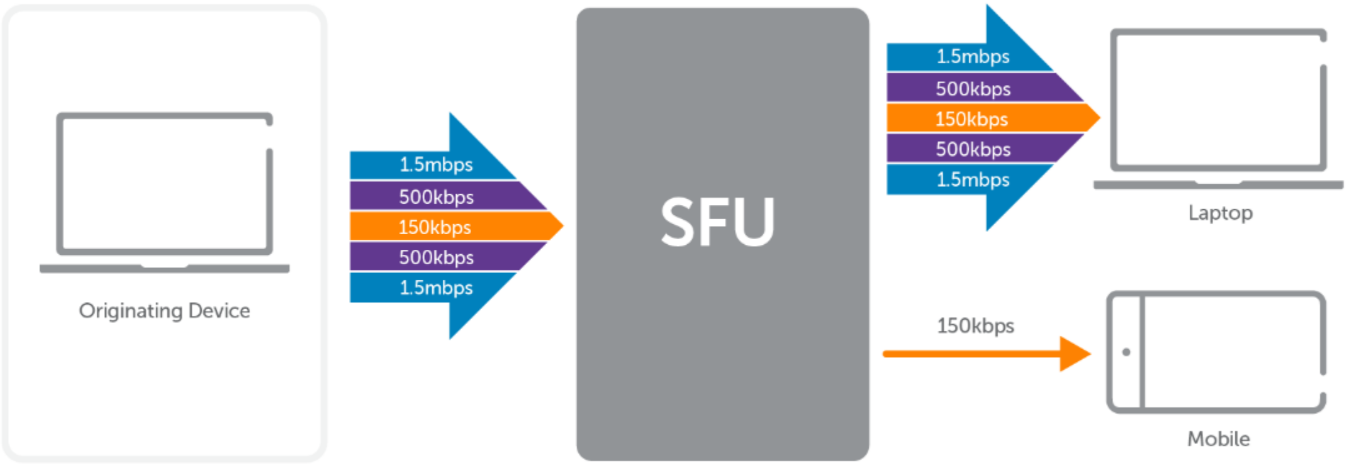
(https://www.cisco.com/c/dam/td-xml/en_us/collaboration/cloud_cmr/pcia_2_0/reports/Troubleshooting_Audio_and_Video_Quality_Using_Webex_Control_Hub.pdf.)

	To support more than four video streams across a distribution link, it is recommended that the bandwidth of the link be set to greater than 2Mbps. Use the API or the Web Admin Interface to set the bandwidth. If using the API, PUT a value for the `peerLinkBitRate` parameter to the API object `/system/configuration/cluster`; the value will be the maximum media bit rate to use on distribution links between Call Bridges in the cluster. Alternatively, using the Web Admin Interface, go to **Configuration > Cluster > Call Bridge identity** and enter the **Peer link bit rate**. (https://www.cisco.com/c/dam/en/us/td/docs/conferencing/ciscoMeetingServer/Deployment_Guide/Version-3-9/Cisco-Meeting-Server-3-9-Scalable-and-Resilient-Deployment.pdf.)
forward the base layer from the video router to at least two of the plurality of video receivers via the internet protocol network, and selectively forward one or more of the set of enhancement layers, but fewer than all of the set of enhancement layers, to at least two of the plurality of video receivers through the internet protocol network based upon the identified bandwidth-limited conditions,	The Accused Instrumentalities forward the base layer from the video router to at least two of the plurality of video receivers via the internet protocol network, and selectively forward one or more of the set of enhancement layers, but fewer than all of the set of enhancement layers, to at least two of the plurality of video receivers through the internet protocol network based upon the identified bandwidth-limited conditions. For example, the AV1 standard discloses forwarding (e.g., selective forwarding using SFU – selective forwarding unit or SFM – selective forwarding middlebox) the base layer (e.g., base layer) to at least two of the plurality of video receivers (e.g., video data receiver such as a video bitstream decoder, etc.) via the internet protocol network (e.g., Internet, etc.) and selectively forwarding (e.g., selective forwarding using SFU – selective forwarding unit or SFM – selective forwarding middlebox) one or more of the set of enhancement layers (e.g., enhancement layers), but fewer than all of the set of enhancement layers (e.g., enhancement layers), to at least two of the plurality of video receivers (e.g., video data receiver such as a video bitstream decoder, etc.) through the internet protocol network (e.g., Internet, etc.) based upon the identified bandwidth-limited conditions (e.g., network condition, available bandwidth condition for a receiving device, etc.). The AV1 standard is a video coding & decoding standard. It discloses that the AV1 standard is developed for video bitstream communication application like Internet related application, streaming, etc. It induces transmission of encoded video bitstream via Internet. It also discloses scaling according to varying bandwidth condition.

	Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) For example, one of the applications of the AV1 standard discloses a selective forwarding unit or middlebox (SFU or SFM) which analyses bandwidth condition between a publisher/transmitter (e.g., video router) and a playback

	device (e.g., video receiver). According to the bandwidth condition, it selectively forwards layers to the playback device. As shown below, it forwards only base layer to devices which are having low bandwidth available and other devices depending on their bandwidth conditions could receive different number of enhancement layers. **How Does SVC Work?** Video Router 1. **Publisher**: Transmits a single media stream (with multiple video layers) to the Selective Forwarding Unit (SFU). 2. **SFU**: Examines the stream and selectively encodes it into high and low bitrate layers. It then distributes these layers based on each playback device's network limitations and available bandwidth. 3. **Playback Devices**: Receive the appropriate layers, ensuring optimal quality for each participant. Video Receiver (https://www.linkedin.com/pulse/scalable-video-coding-svc-sudeep-kumar-nofdc.)

	The SFU then sends the adjusted stream along to the appropriate **playback device**. Depending on the available bandwidth and other limitations of the target devices, some could receive a lower quality stream while others get the whole onion (so to speak).  (https://www.wowza.com/blog/scalable-video-coding-for-webrtc.) **Selective Forwarding Middlebox (SFM)** A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant.

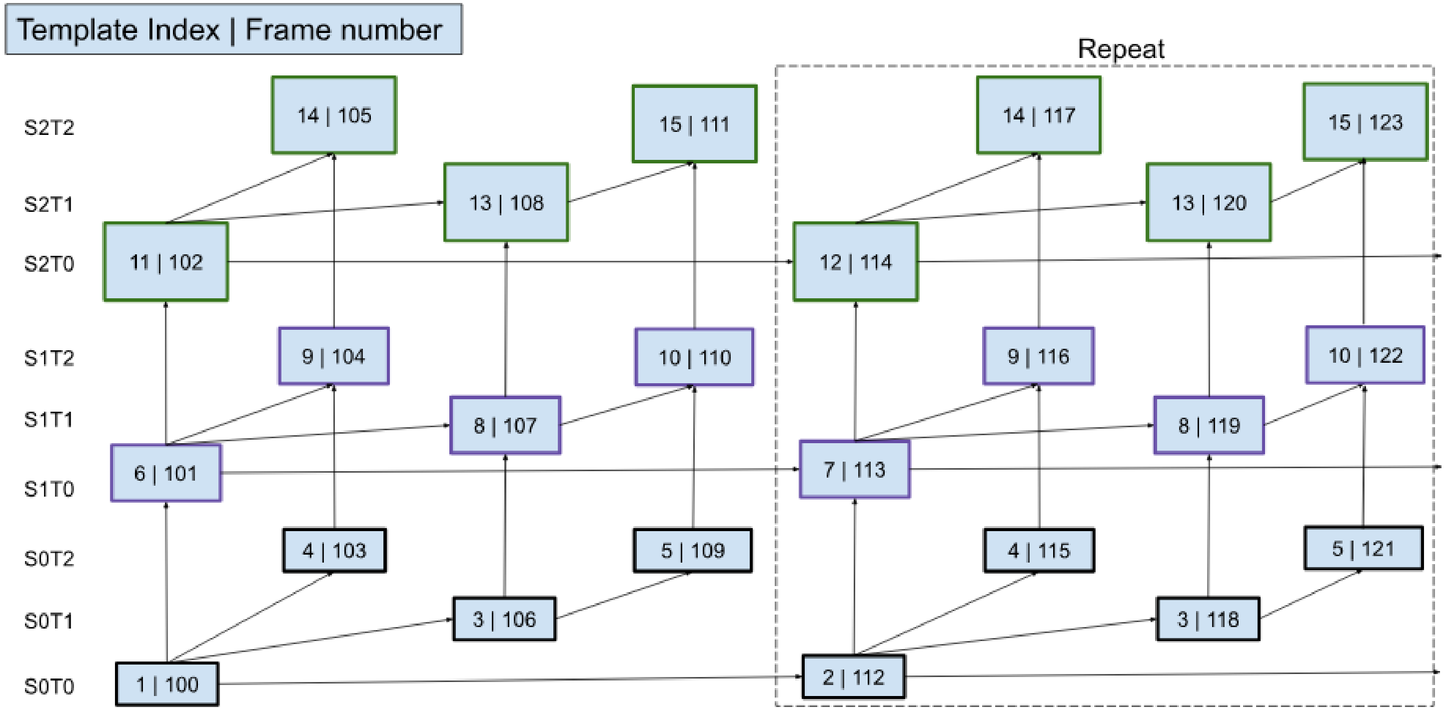
	Dependency Descriptor RTP Header Extension **A.1 Introduction** This appendix describes the Dependency Descriptor (DD) RTP Header extension. The DD is used for conveying dependency information about individual video frames in a scalable video stream. The DD includes provisions for both temporal and spatial scalability. In the DD, the smallest unit for which dependencies are described is an RTP frame. An RTP frame contains one complete coded video frame and may also contain additional information (e.g., metadata). Each RTP frame is identified by a frame_number. When spatial scalability is used, there may be multiple RTP frames produced for the same time instant. Further, this specification allows for the transmission of an RTP frame over multiple packets. RTP frame aggregation is explicitly disallowed. Hereinafter, RTP frame will be referred to as frame. The DD uses the concept of Decode targets, each of which represents a subset of a scalable video stream necessary to decode the stream at a certain temporal and spatial fidelity. A frame may be associated with several Decode targets. This concept is used to facilitate selective forwarding, as is done by a Selective Forwarding Middlebox (SFM). Typically an SFM would select one Decode target for each endpoint, and forward all frames associated with that target.

	# 6 MANE and SFM Behavior If a packet contains an OBU with an OBU extension header then the entire packet is considered associated with the layer identified by the temporal_id and spatial_id combination that are indicated in the extension header. If a packet does not contain any OBU with an OBU extension header, then it is considered to be associated with all operating points. The general function of a MANE or SFM is to selectively forward packets to receivers. To make forwarding decisions a MANE may inspect the media payload, so that it may need to be able to parse the AV1 bitstream and if so, cannot function when end-to-end encryption is enabled. An SFM does not parse the AV1 bitstream and therefore needs to obtain the information relevant to selective forwarding by other means, such as the Dependency Descriptor described in Appendix A. In addition to enabling bitstream-independent forwarding and support for end-to-end encryption, the Dependency Descriptor also enables forwarding where the metadata OBU provided in the AV1 bitstream is not sufficient to express the structure of the stream. ### 6.1.1 Example In this example, it is desired to send three simulcast encodings, each containing three temporal layers. When sending all encodings on a single SSRC, scalability mode 'S3T3' would be indicated within the scalability metadata OBU, and the Dependency Descriptor describes the dependency structure of all encodings. When sending each simulcast encoding on a distinct SSRC, the scalability mode 'L1T3' would be indicated within the scalability metadata OBU of each bitstream, and the Dependency Descriptor in each stream describes only the dependency structure for that individual encoding. A distinct spatial_id (e.g. 0, 1, 2) could be used for each stream (if a single AV1 encoder is used to produce the three simulcast encodings), but if distinct AV1 encoders are used, the spatial_id values may not be distinct.

	Indication	Description	SFM behavior
DT0	Not present	F5 is not associated with DT0.	Do not forward F5 to a DT0 client.
DT1	Discardable	No frame in DT1 will reference F5.	Should forward F5 to a DT1 client, but may discard it, e.g., when bandwidth is low.
DT2	Switch	If it is possible to decode F5, then all future frames in DT2 are also decodable.	Forward F5 to the DT2 client. In addition, the SFM can rely on the fact that if the F5 frame is decodable for a particular client then that client may be switched to DT2.
DT3	Required	Future frames in DT3 reference F5.	Forward F5 to a DT3 client. No additional decisions can be made.

A.10.2.2 L3T3 Full SVC

This example uses three Chains. Chain 0 includes frames with spatial ID equal to 0 and temporal ID equal to 0. Chain 1 includes frames with spatial ID equal to 0 or 1 and temporal ID equal to 0. Chain 2 includes all frames with temporal ID equal to 0.



Idx	S	T	Fdiffs	Chains			Decode Targets								
				0	1	2	HD30 fps	HD15 fps	HD7.5 fps	VGA30 fps	VGA15 fps	VGA7.5fps	QVGA30 fps	QVGA15 fps	QVGA7.5 fps
1	0	0		0	0	0	S	S	S	S	S	S	S	S	S
2	0	0	12	12	11	10	R	R	R	R	R	R	S	S	S
3	0	1	6	6	5	4	R	R	-	R	R	-	S	D	-
4	0	2	3	3	2	1	R	-	-	R	-	-	D	-	-
5	0	2	3	9	8	7	R	-	-	R	-	-	D	-	-
6	1	0	1	1	1	1	S	S	S	S	S	S	-	-	-
7	1	0	12,1	1	1	1	R	R	R	S	S	S	-	-	-
8	1	1	6,1	7	6	5	R	R	-	S	D	-	-	-	-
9	1	2	3,1	4	3	2	R	-	-	D	-	-	-	-	-
10	1	2	3,1	10	9	8	R	-	-	D	-	-	-	-	-
11	2	0	1	2	1	1	S	S	S	-	-	-	-	-	-
12	2	0	12,1	2	1	1	S	S	S	-	-	-	-	-	-
13	2	1	6,1	8	7	6	S	D	-	-	-	-	-	-	-
14	2	2	3,1	5	4	3	D	-	-	-	-	-	-	-	-
15	2	2	3,1	11	10	9	D	-	-	-	-	-	-	-	-
decode_target_protected_by							2	2	2	1	1	1	0	0	0

(https://aomediacodec.github.io/av1-rtp-spec/.)

	Base layer The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **Enhancement layer** A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)* 6.2.3. OBU extension header semantics **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)*

	Note: The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). Coded video data of a temporal level with temporal_id T and spatial level with spatial_id S are only allowed to reference previously coded video data of temporal_id T' and spatial_id S', where T' <= T and S' <= S. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* If a coded video sequence contains at least one enhancement layer (OBUs with spatial_id greater than 0 or temporal_id greater than 0) then all frame headers and tile group OBUs associated with base (spatial_id equals 0 and temporal_id equals 0) and enhancement layer (spatial_id greater than 0 or temporal_id greater than 0) data must include the OBU extension header. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 205 of 669)*

Table 4. Webex Meetings Bandwidth per Resolution Table

Layer	Bandwidth Range
90p active thumbnail (each)	~60-100 kb/s
180p main video	125-200 kb/s
360p main video	470-640 kb/s
720p main video	900k-1.5 mb/s
Content sharing (sharpness, 1080p/5)	120k – 1.3 mb/s
Content sharing (motion, 720p/30)	900k – 2.5 mb/s

Webex Meetings Desktop App Bandwidth Controls

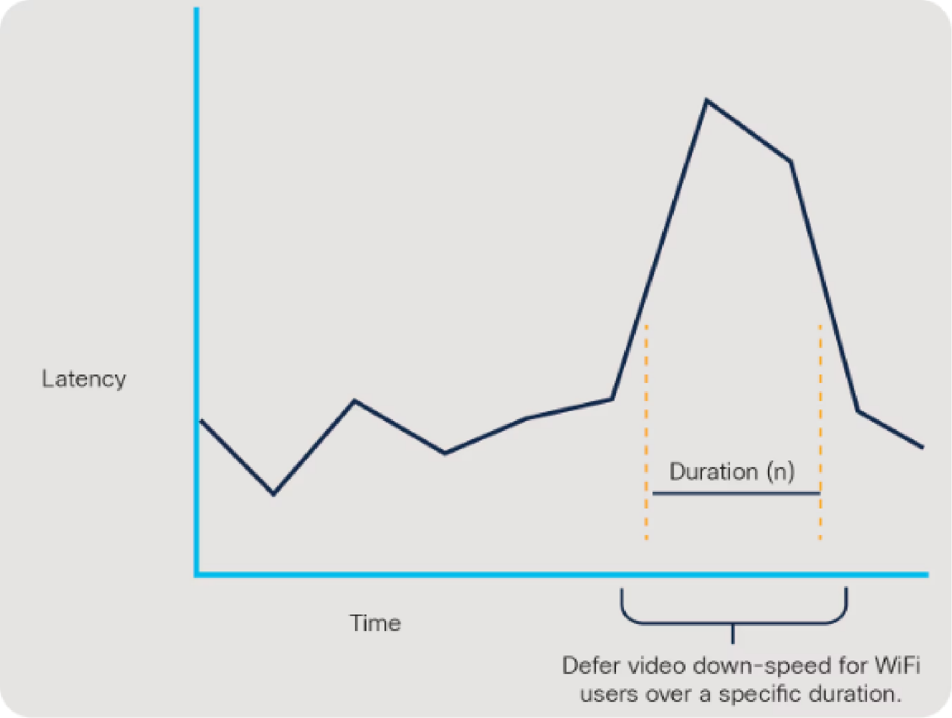
Webex administrators have 2 key controls to help control bandwidth as used by clients that connect to Webex meetings should they choose to. Namely, you can cap the meeting layouts at either 360p as the max available resolution, or to enable 720p layers. Whether your site is administered on Webex Control Hub or Webex Site Administrator, the following controls are available in Configuration > Common Site Settings > Options:

☐ Turn on high-quality video (360p) *(Meetings, Training, Events and Support)*

☐ Turn on high-definition video (720p) *(Meetings, Training and Events)*

Figure 5.

Webex Meetings Desktop App Bandwidth Controls

	Webex Media Improvements The following are media improvements that have occurred in releases from 40.7 – 40.10 **Improved Video Experience over WIFI (Defer video down-speeding):** WIFI environments present a specific set of challenges and through testing we noticed the algorithms were too quick to drop resolution/bandwidth in certain WIFI environments. In 40.7 the algorithms have been updated to 'defer the down-speeding" of video for brief WIFI specific network dropouts. This resulted in improved user experience over WIFI. See figure5 for an illustration of this. When latency spikes for a specific duration the improved algorithms can detect that this is a WIFI specific event and not thus delay the down-speeding of the video resolution until that duration (n) is exceeded in which case would infer a different network event increasing latency.  **Figure 6.** Deferred Video Down-speeding

	Video Super Scaling is a technique used for users on sub-optimal networks. We provide 720p like experiences to users who are experiencing network impairments resulting in 360p connections to our servers for their meeting. Our updated scaling algorithm provides enhanced video quality when this degradation occurs, making network impairment downgrades less quality impacting. If a 360p image is simply scaled geometrically to a 720p resolution, the quality has not improved, but our technique involves migrating video post-processing tasks to the client for some video flows versus doing it in the cloud, resulting in being able to keep more precious bits of information to reconstruct the video at a given bandwidth. In a meeting as shown in figure 8 and 9, the bandwidth requirements are as follows. Assuming the receiver is not bandwidth restricted, Webex app will adapt to available network conditions. Table 8 outlines the bandwidth utilization. Figure 13 illustrates the Active Speaker layout which is the default meeting layout for Webex Teams App. **Table 8. Cisco Webex Teams Bandwidth Utilization** (see table below) (https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings/white_paper_c11-691351.html.)

Meeting Scenario	Main Video+Audio (no content sharing)	Main Video+Audio (content sharing)
Webex Teams 1:1 Call	1.5 mb/s peak, 1.3 mb/s avg	1.5 mb/s peak, 1.3 mb/s avg
Webex Teams Meeting (fig 13 layout), 6 participants	2.3 mb/s peak, 1.7 mb/s avg	2.9/ mb/s peak, 2.0 mb/s avg
Webex Teams Meeting (fig 14 layout), 6 participants	2.5 mb/s peak, 1.8 mb/s avg	3.5 mb/s peak, 2.3 mb/s avg

and wherein the video router transmits the layered video data stream according to an internet protocol;	The Accused Instrumentalities include a video router, wherein the video router transmits the layered video data stream according to an internet protocol. For example, the AV1 standard discloses the method such that the layered video data stream (e.g., scalable video bitstream, etc.) is transmitted according to an internet protocol (e.g., Internet, etc.). Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.) **Selective Forwarding Middlebox (SFM)** A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant. (https://aomediacodec.github.io/av1-rtp-spec/.)

<table>
<tr><td></td><td>

AV1 Features

ROYALTY-FREE

Interoperable and open

UBIQUITOUS

Scales to any modern device at any bandwith

FLEXIBLE

For use in both commercial and non-commercial content, including user-generated content

30% BETTER COMPRESSION *

Uses less data while delivering 4k UHD video and beyond when compared to alternatives

OPTIMIZED

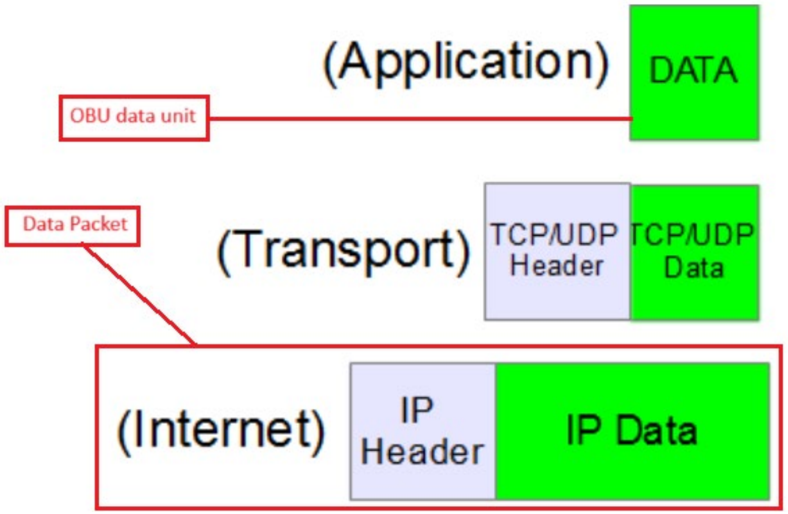
Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services

(https://aomedia.org/av1-features/.)

</td></tr>
<tr><td>wherein each layer of the layered video data stream comprises data packets,</td><td>The Accused Instrumentalities include a video router, wherein each layer of the layered video data stream comprises data packets, each of which is encoded with a sequence number and a layer identifier, and wherein the layer identifier for each data packet is based upon a layer to which the packet belongs.</td></tr>
</table>

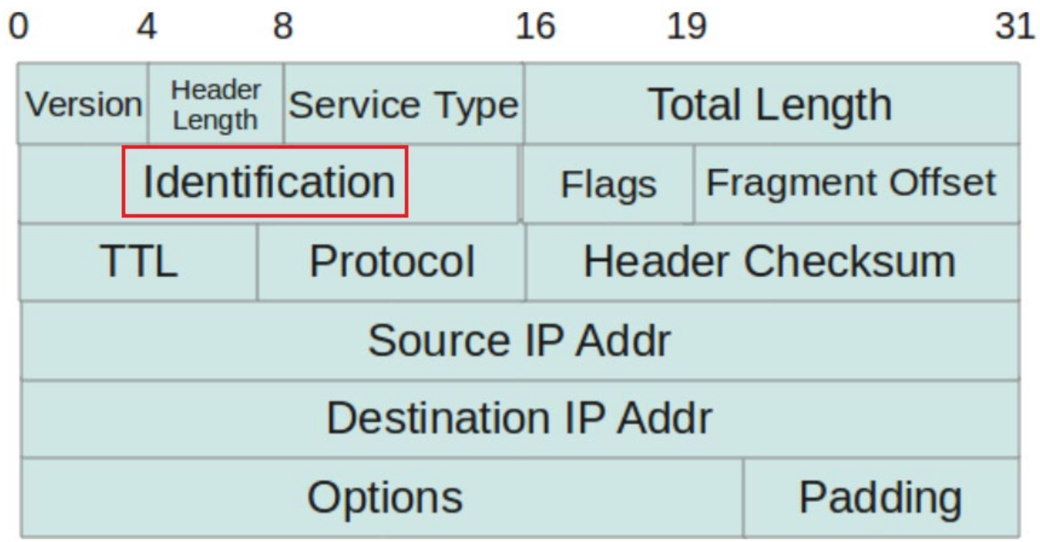
each of which is encoded with a sequence number and a layer identifier, and wherein the layer identifier for each data packet is based upon a layer to which the packet belongs	For example, the AV1 standard discloses the method such that each layer of the layered video data stream comprises data packets (e.g., IP packets of video bitstream data units) each of which is encoded with a sequence number (e.g., identification number value of IP packet, etc.) and a layer identifier (e.g., a layer identifier such as base layer, enhancement layer, etc.) and wherein the layer identifier (e.g., the layer identifier such as base layer, enhancement layer, etc.) for each data packet (e.g., IP packets of video bitstream data units) is based upon a layer (e.g., layer such as base layer, enhancement layer, etc.) to which the packet belongs. As shown below, the AV1 standard discloses an encoded video data bitstream using scalable video coding in a sequence of OBUs i.e., open bitstream unit. The OBU data units are transmitted in packetized format over Internet. These packets are governed by IP protocol and communicated as an IP packet. An IP packet comprises a header part and a payload/data part. The header part of the IP packet comprises an identification field which denotes a sequence number of IP packets i.e., data packets, transmitted.

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) **Bitstream** The sequence of bits generated by encoding a sequence of frames. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)*

	OBU All structures are packetized in "Open Bitstream Units" or OBUs. Each OBU has a header, which provides identifying information for the contained data (payload). *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page – 4 of 669)* **Open Bitstream Unit (OBU)** The smallest bitstream data framing unit in AV1. All AV1 bitstream structures are packetized in OBUs. (https://aomediacodec.github.io/av1-rtp-spec/#3-media-format-description.)  (https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

	These are a set of standard rules that allows different types of computers to communicate with each other. The IP protocol ensures that each computer that is connected to the Internet is having a specific serial number called the IP address. TCP specifies how data is exchanged over the internet and how it should be broken into IP packets. It also makes sure that the packets have information about the source of the message data, the destination of the message data, the sequence in which the message data should be re-assembled, and checks if the message has been sent correctly to the specific destination. The TCP is also known as a connection-oriented protocol. (https://www.geeksforgeeks.org/types-of-internet-protocols/.)

IP Header



(https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

- ***Identification(16 bits)***: This field is used for uniquely identifying the IP datagrams. This value is incremented every-time an IP datagram is sent from source to the destination. This field comes in handy while reassembly of fragmented IP data grams.

(https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

Further, the OBU data unit comprises a metadata syntax which discloses scalability corresponding to the OBU. It discloses three types of scalabilities i.e., Spatial scalability, Temporal scalability and Quality scalability. These

	scalabilities define a spatial layer having a corresponding spatial_id and a temporal layer having a corresponding temporal_id. Further, the AV1 standard discloses deriving a layered coded bitstream of base layer and enhancement layers using scalable video coding. It discloses a base layer having both spatial_id and temporal_id equal to zero and enhancement layers with at least one of spatial_id or temporal_id values greater than zero. **Bitstream** The sequence of bits generated by encoding a sequence of frames. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **5.8.1. General metadata OBU syntax**

metadata_obu( ) {	Type
metadata_type	leb128()
if (metadata_type == METADATA_TYPE_ITUT_T35)	
metadata_itut_t35( )	
else if (metadata_type == METADATA_TYPE_HDR_CLL)	
metadata_hdr_cll( )	
else if (metadata_type == METADATA_TYPE_HDR_MDCV)	
metadata_hdr_mdcv( )	
else if (metadata_type == METADATA_TYPE_SCALABILITY)	
metadata_scalability( )	
else if (metadata_type == METADATA_TYPE_TIMECODE)	
metadata_timecode( )	
}	

Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 34 of 669)

	5.8.5. Metadata scalability syntax
	
	`metadata_scalability( ) {` — Type
	    `scalability_mode_idc` — `f(8)`
	    `if ( scalability_mode_idc == SCALABILITY_SS )`
	        `scalability_structure( )`
	`}`
	
	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

5.8.6. Scalability structure syntax

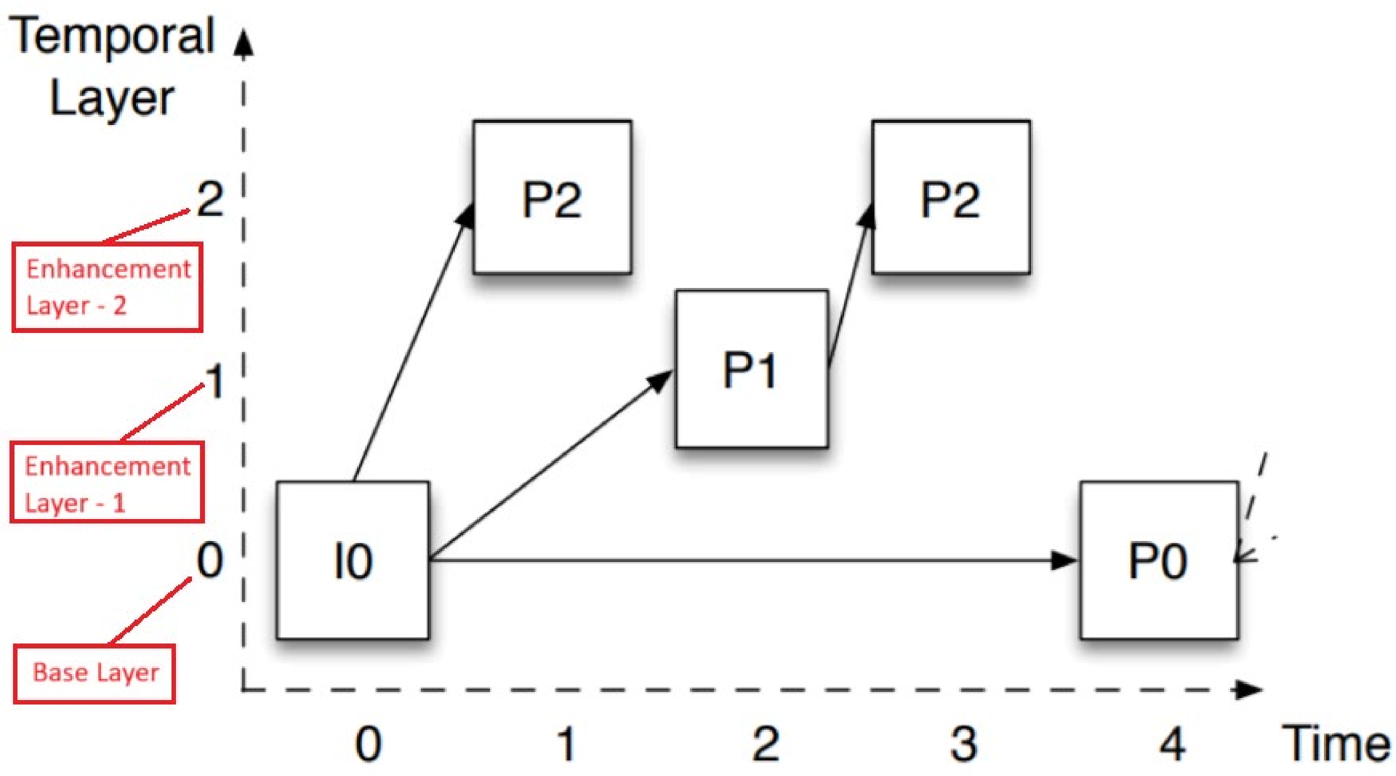
scalability_structure( ) {	Type
spatial_layers_cnt_minus_1	f(2)
spatial_layer_dimensions_present_flag	f(1)
spatial_layer_description_present_flag	f(1)
temporal_group_description_present_flag	f(1)
scalability_structure_reserved_3bits	f(3)
if (spatial_layer_dimensions_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1 ; i++) {	
spatial_layer_max_width[i]	f(16)
spatial_layer_max_height[i]	f(16)
}	
}	
if (spatial_layer_description_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1; i++)	
spatial_layer_ref_id[i]	f(8)
}	
if (temporal_group_description_present_flag) {	
temporal_group_size	f(8)
for (i = 0; i < temporal_group_size; i++) {	
temporal_group_temporal_id[i]	f(3)
temporal_group_temporal_switching_up_point_flag[i]	f(1)
temporal_group_spatial_switching_up_point_flag[i]	f(1)
temporal_group_ref_cnt[i]	f(3)
for (j = 0; j < temporal_group_ref_cnt[i]; j++) {	
temporal_group_ref_pic_diff[i][j]	f(8)
}	

Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

spatial_layer_ref_id[i] specifies the spatial_id value of the frame within the current temporal unit that the frame of layer i uses for reference. If no frame within the current temporal unit is used for reference the value must be equal to 255.

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669) Note that for a given picture, all frames follow the same inter-picture temporal dependency structure. However, the frame rate of each layer can be different from each other. The specified dependency structure in the scalability structure data must be for the highest frame rate layer. **temporal_group_temporal_id[i]** specifies the temporal_id value for the i-th picture in the temporal group. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669)* **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)* **Base layer** The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **Enhancement layer** A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)*

	Note: The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). Coded video data of a temporal level with temporal_id T and spatial level with spatial_id S are only allowed to reference previously coded video data of temporal_id T' and spatial_id S', where T' <= T and S' <= S. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* **Note:** Examples are given for non-scalable cases, but the constraints also apply to each operating point of a scalable stream. For example, consider a 60fps spatial scalable stream with a base layer at 960x540, and an enhancement layer at 1920x1080. The operating point containing just the base layer may be labelled as level 3.0, while the operating point containing both the base and enhancement layer may be labelled as level 4.1. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 641 of 669)*

	 *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 126 of 669)*

Dependency Descriptor RTP Header Extension

A.1 Introduction

This appendix describes the Dependency Descriptor (DD) RTP Header extension. The DD is used for conveying dependency information about individual video frames in a scalable video stream. The DD includes provisions for both temporal and spatial scalability.

In the DD, the smallest unit for which dependencies are described is an RTP frame. An RTP frame contains one complete coded video frame and may also contain additional information (e.g., metadata). Each RTP frame is identified by a frame_number. When spatial scalability is used, there may be multiple RTP frames produced for the same time instant. Further, this specification allows for the transmission of an RTP frame over multiple packets. RTP frame aggregation is explicitly disallowed. Hereinafter, RTP frame will be referred to as frame.

The DD uses the concept of Decode targets, each of which represents a subset of a scalable video stream necessary to decode the stream at a certain temporal and spatial fidelity. A frame may be associated with several Decode targets. This concept is used to facilitate selective forwarding, as is done by a Selective Forwarding Middlebox (SFM). Typically an SFM would select one Decode target for each endpoint, and forward all frames associated with that target.

A.8 Dependency Descriptor Format

To facilitate the work of selectively forwarding portions of a scalable video bitstream, as is done by an SFM, for each packet, the following information is made available (even though not all elements are present in every packet).

- spatial ID
- temporal ID
- DTIs
- frame_number of the current frame
- frame_number of each of the Referred frames
- frame_number of last frame in each Chain

6 MANE and SFM Behavior

If a packet contains an OBU with an OBU extension header then the entire packet is considered associated with the layer identified by the temporal_id and spatial_id combination that are indicated in the extension header. If a packet does not contain any OBU with an OBU extension header, then it is considered to be associated with all operating points.

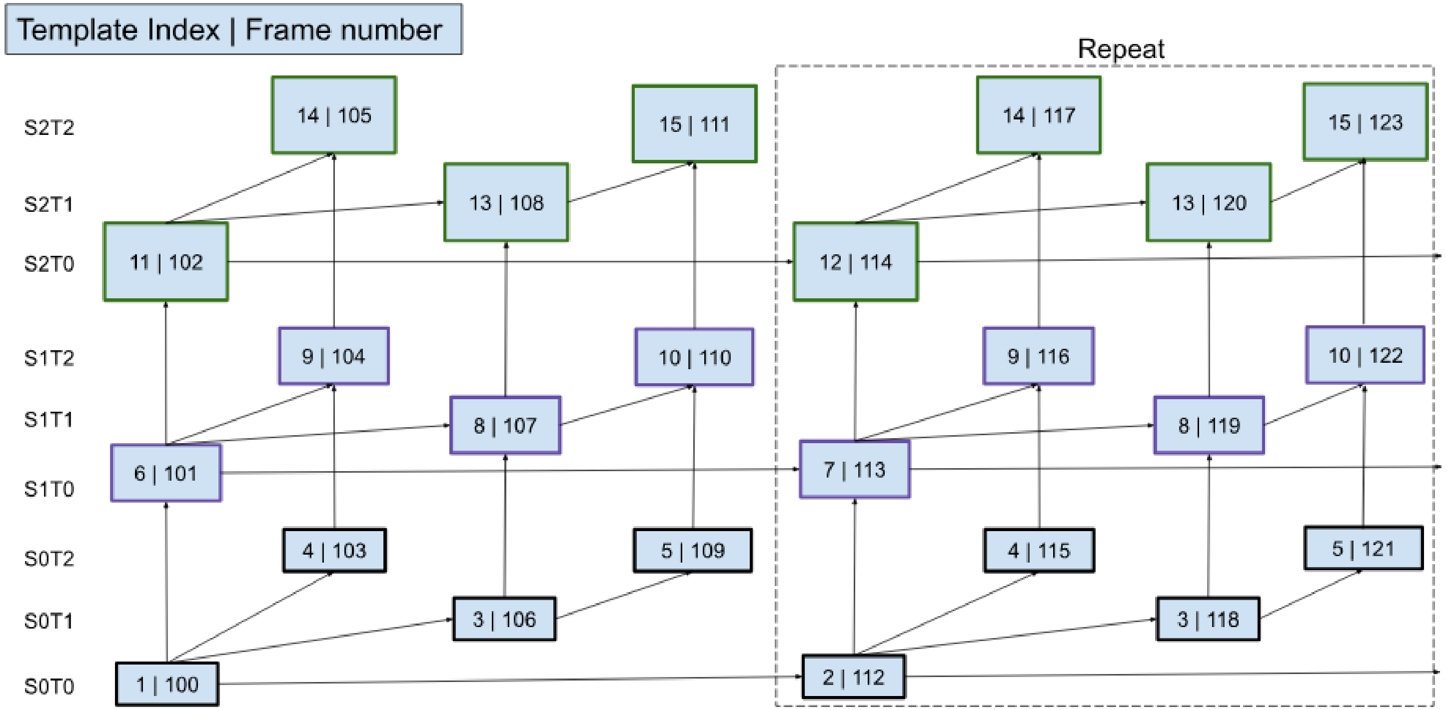
The general function of a MANE or SFM is to selectively forward packets to receivers. To make forwarding decisions a MANE may inspect the media payload, so that it may need to be able to parse the AV1 bitstream and if so, cannot function when end-to-end encryption is enabled. An SFM does not parse the AV1 bitstream and therefore needs to obtain the information relevant to selective forwarding by other means, such as the Dependency Descriptor described in Appendix A. In addition to enabling bitstream-independent forwarding and support for end-to-end encryption, the Dependency Descriptor also enables forwarding where the metadata OBU provided in the AV1 bitstream is not sufficient to express the structure of the stream.

6.1.1 Example

In this example, it is desired to send three simulcast encodings, each containing three temporal layers. When sending all encodings on a single SSRC, scalability mode 'S3T3' would be indicated within the scalability metadata OBU, and the Dependency Descriptor describes the dependency structure of all encodings.

When sending each simulcast encoding on a distinct SSRC, the scalability mode 'L1T3' would be indicated within the scalability metadata OBU of each bitstream, and the Dependency Descriptor in each stream describes only the dependency structure for that individual encoding. A distinct spatial_id (e.g. 0, 1, 2) could be used for each stream (if a single AV1 encoder is used to produce the three simulcast encodings), but if distinct AV1 encoders are used, the spatial_id values may not be distinct.

		Indication	Description	SFM behavior
	DT0	Not present	F5 is not associated with DT0.	Do not forward F5 to a DT0 client.
	DT1	Discardable	No frame in DT1 will reference F5.	Should forward F5 to a DT1 client, but may discard it, e.g., when bandwidth is low.
	DT2	Switch	If it is possible to decode F5, then all future frames in DT2 are also decodable.	Forward F5 to the DT2 client. In addition, the SFM can rely on the fact that if the F5 frame is decodable for a particular client then that client may be switched to DT2.
	DT3	Required	Future frames in DT3 reference F5.	Forward F5 to a DT3 client. No additional decisions can be made.

	A.10.2.2 L3T3 Full SVC This example uses three Chains. Chain 0 includes frames with spatial ID equal to 0 and temporal ID equal to 0. Chain 1 includes frames with spatial ID equal to 0 or 1 and temporal ID equal to 0. Chain 2 includes all frames with temporal ID equal to 0. 

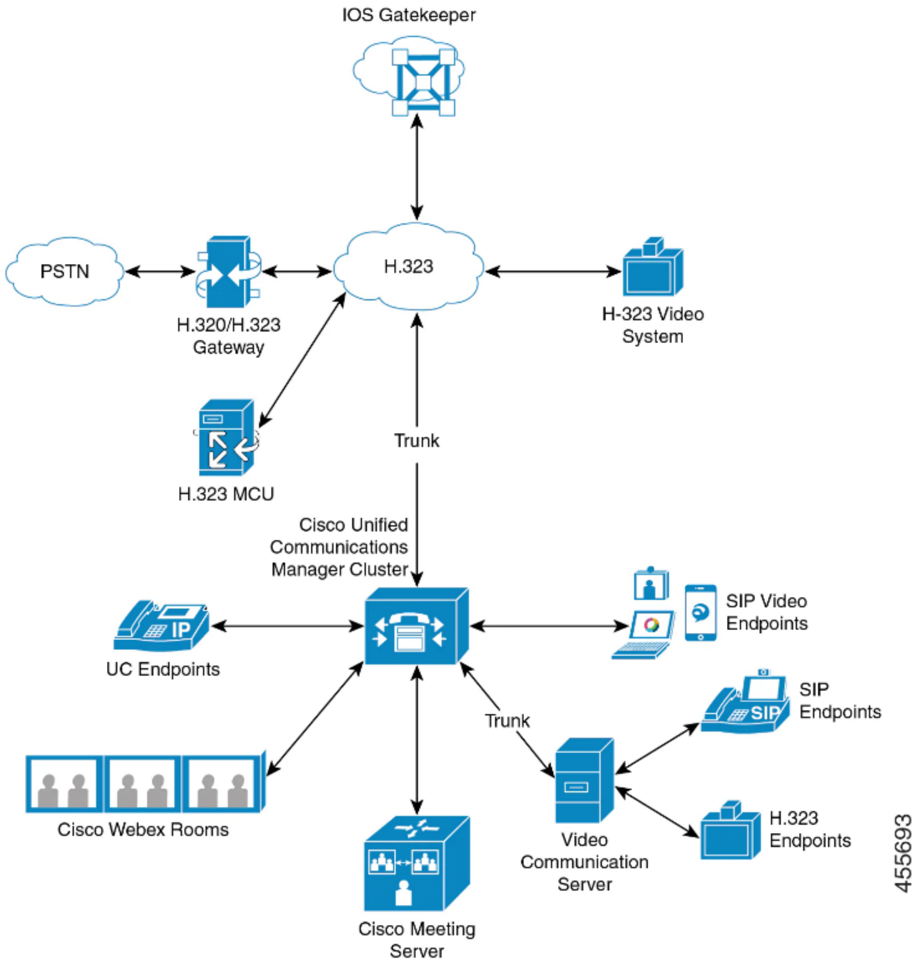
Idx	S	T	Fdiffs	Chains			Decode Targets								
				0	1	2	HD30 fps	HD15 fps	HD7.5 fps	VGA30 fps	VGA15 fps	VGA7.5fps	QVGA30 fps	QVGA15 fps	QVGA7.5 fps
1	0	0		0	0	0	S	S	S	S	S	S	S	S	S
2	0	0	12	12	11	10	R	R	R	R	R	R	S	S	S
3	0	1	6	6	5	4	R	R	-	R	R	-	S	D	-
4	0	2	3	3	2	1	R	-	-	R	-	-	D	-	-
5	0	2	3	9	8	7	R	-	-	R	-	-	D	-	-
6	1	0	1	1	1	1	S	S	S	S	S	S	-	-	-
7	1	0	12,1	1	1	1	R	R	R	S	S	S	-	-	-
8	1	1	6,1	7	6	5	R	R	-	S	D	-	-	-	-
9	1	2	3,1	4	3	2	R	-	-	D	-	-	-	-	-
10	1	2	3,1	10	9	8	R	-	-	D	-	-	-	-	-
11	2	0	1	2	1	1	S	S	S	-	-	-	-	-	-
12	2	0	12,1	2	1	1	S	S	S	-	-	-	-	-	-
13	2	1	6,1	8	7	6	S	D	-	-	-	-	-	-	-
14	2	2	3,1	5	4	3	D	-	-	-	-	-	-	-	-
15	2	2	3,1	11	10	9	D	-	-	-	-	-	-	-	-
decode_target_protected_by							2	2	2	1	1	1	0	0	0

(https://aomediacodec.github.io/av1-rtp-spec/.)

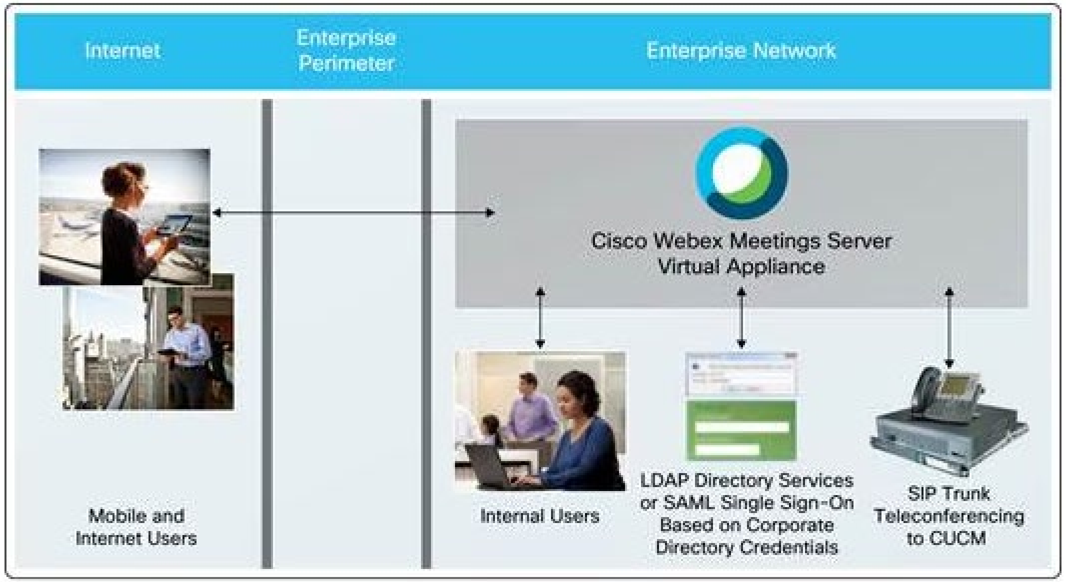
6. A method for transmitting video signals, the method comprising:	The Accused Instrumentalities perform a method for transmitting video signals. Cisco performs the claimed method, for example, by using the Accused Instrumentalities in its provision of monthly subscription services to Webex and Cisco Meeting Server. For example:

Video Network

The following illustration provides an example of a video network that uses a single Unified Communications Manager cluster. In a successful video network, any endpoint can call any other endpoint. Video availability only exists if both endpoints are video-enabled. Video capabilities extend across trunks.



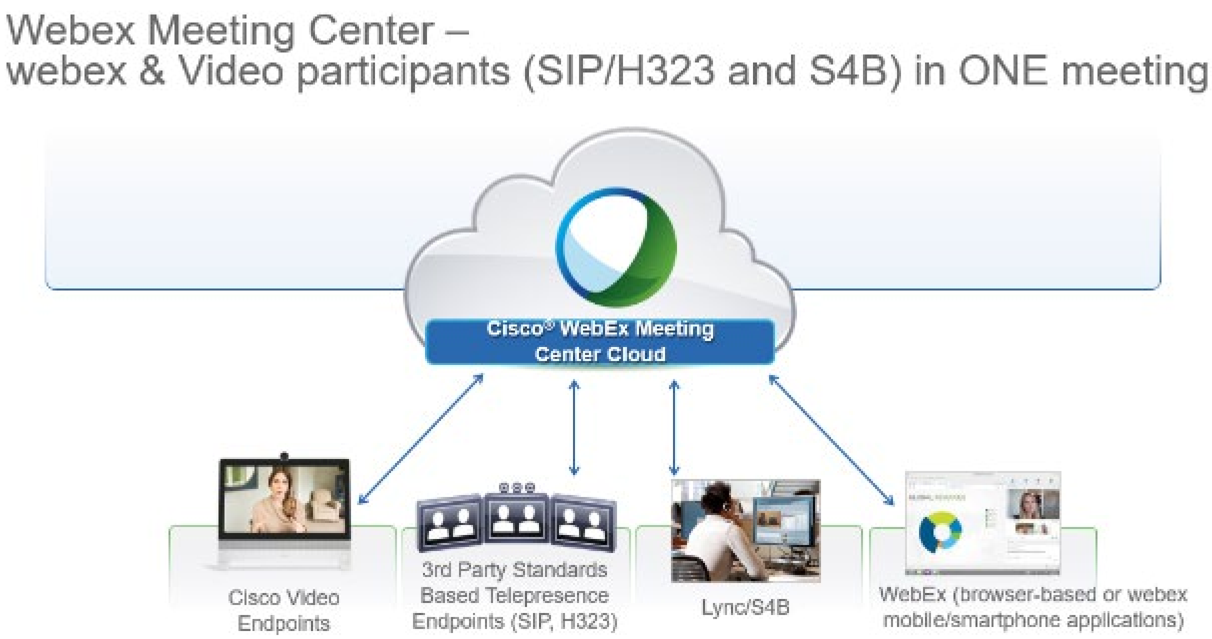
	The Cisco video conference portfolio comprises the following video bridges: • Cisco TelePresence MCU series • Webex Meeting Server (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/14SU2/cucm_b_feature-configuration-guide-for-cisco14su2/cucm_m_video-telephony.html.) **Product Overview** Cisco Webex Meetings Server is a virtualized, software-based solution that runs on Cisco Unified Computing System™ (Cisco UCS®) servers and VMware. It uses virtual appliance technology for rapid turn-up of services to end users. With Cisco Webex Meetings Server, there are two options for enabling mobile users to more securely access Cisco Webex conferences without going through a VPN. The first option is to deploy reverse proxy (or edge servers) in the enterprise perimeter (or DMZ). The second option, shown in Figure 1, is to deploy the reverse proxy servers behind your internal firewall, thus eliminating all DMZ components and related information security concerns.

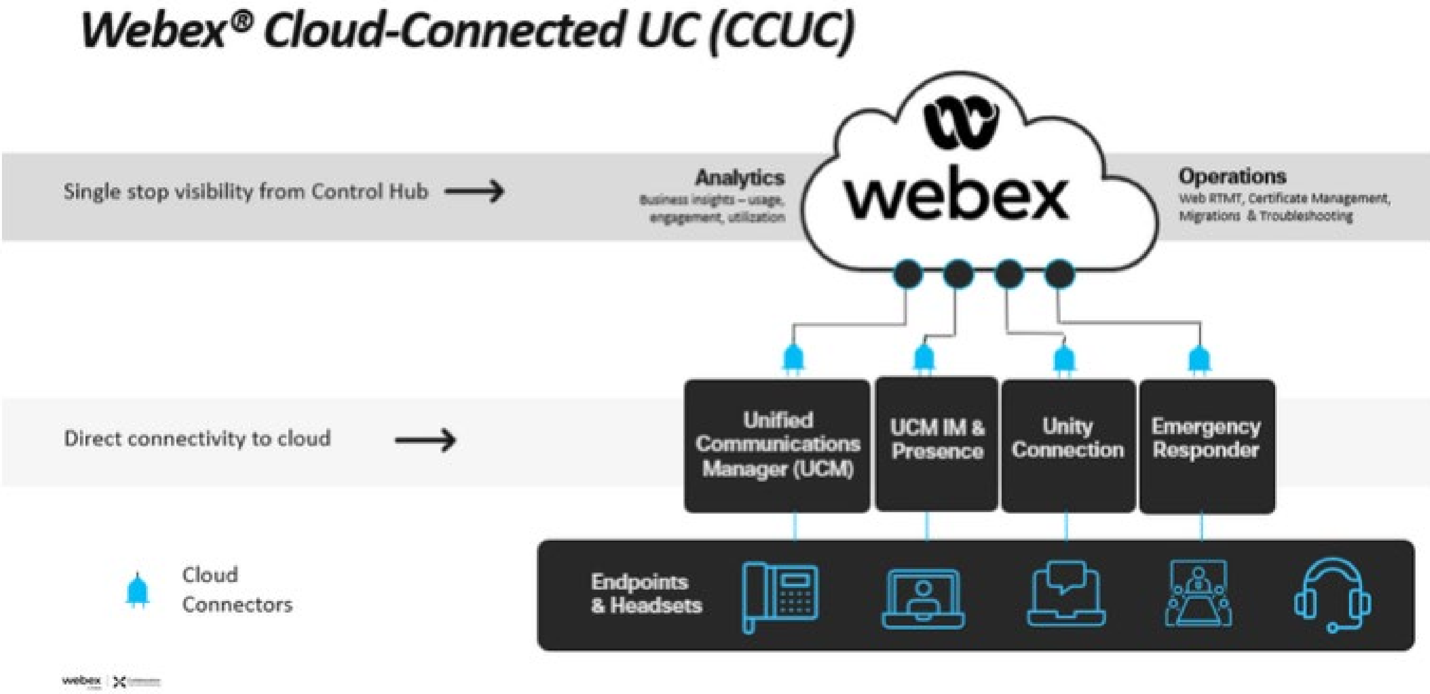
	 Optimized for 100% Secure, Behind-the-Firewall VPN-Less Access That Integrates with Your Corporate User Management and UC Infrastructure **Figure 1.** Full Deployment of Cisco Webex Meetings Server Behind a Firewall

System Requirements

Cisco Webex Meetings Server is compatible with Cisco UCS servers that meet or exceed the specifications presented in this section.

Module	Requirements
Host server	Cisco UCS C-Series rack server or equivalent B-Series blade server
Network interfaces	• Minimum 1 physical Network Interface Card (NIC) for a nonredundant configuration • Redundant configurations must have all NIC interfaces duplicated and connected to an independent switching fabric
Internal storage (direct attached storage [DAS]) for ESXi hosts where internal machines are deployed	Minimum of 4 drives in a RAID-10 or RAID-5 configuration
Internal storage (DAS)for ESXi hosts where Internet Reverse Proxy (IRP) virtual machines are deployed	Minimum of 2 drives in a RAID-1 configuration
SAN storage	Can be used as a substitute for DAS
Network-Attached Storage (NAS)	Can be used as a substitute for DAS or SAN
Hypervisor	• ESXi versions and vSphere licenses • 1 VMware license per processor socket
Email server	• Fully Qualified Domain Name (FQDN) of the mail server that the system uses to send emails • Port number: Default value of the Simple Mail Transfer Protocol (SMTP) port number is 25 or 465 (secure SMTP port number)

	(https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings-server/datasheet-c78-717754.html.)  (https://community.cisco.com/t5/collaboration-knowledge-base/cisco-webex-meetings-overview/ta-p/3648888.)

	 (https://community.cisco.com/t5/collaboration-knowledge-base/webex-cloud-connected-uc-connect-your-cucm-to-webex-cloud-for/ta-p/4172103.)

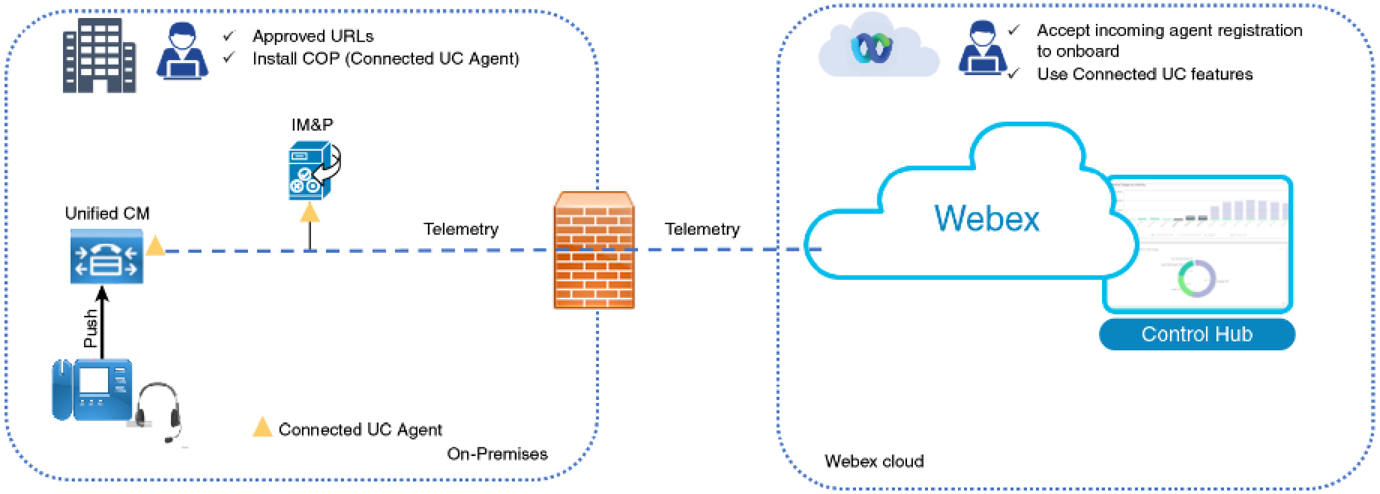
	 *Cloud-Connected UC Deployment* (https://help.webex.com/en-us/article/jv0u1db/Webex-Cloud-Connected-UC-Overview.)

Figure 11-14 Cisco WebEx Collaboration Cloud Architecture

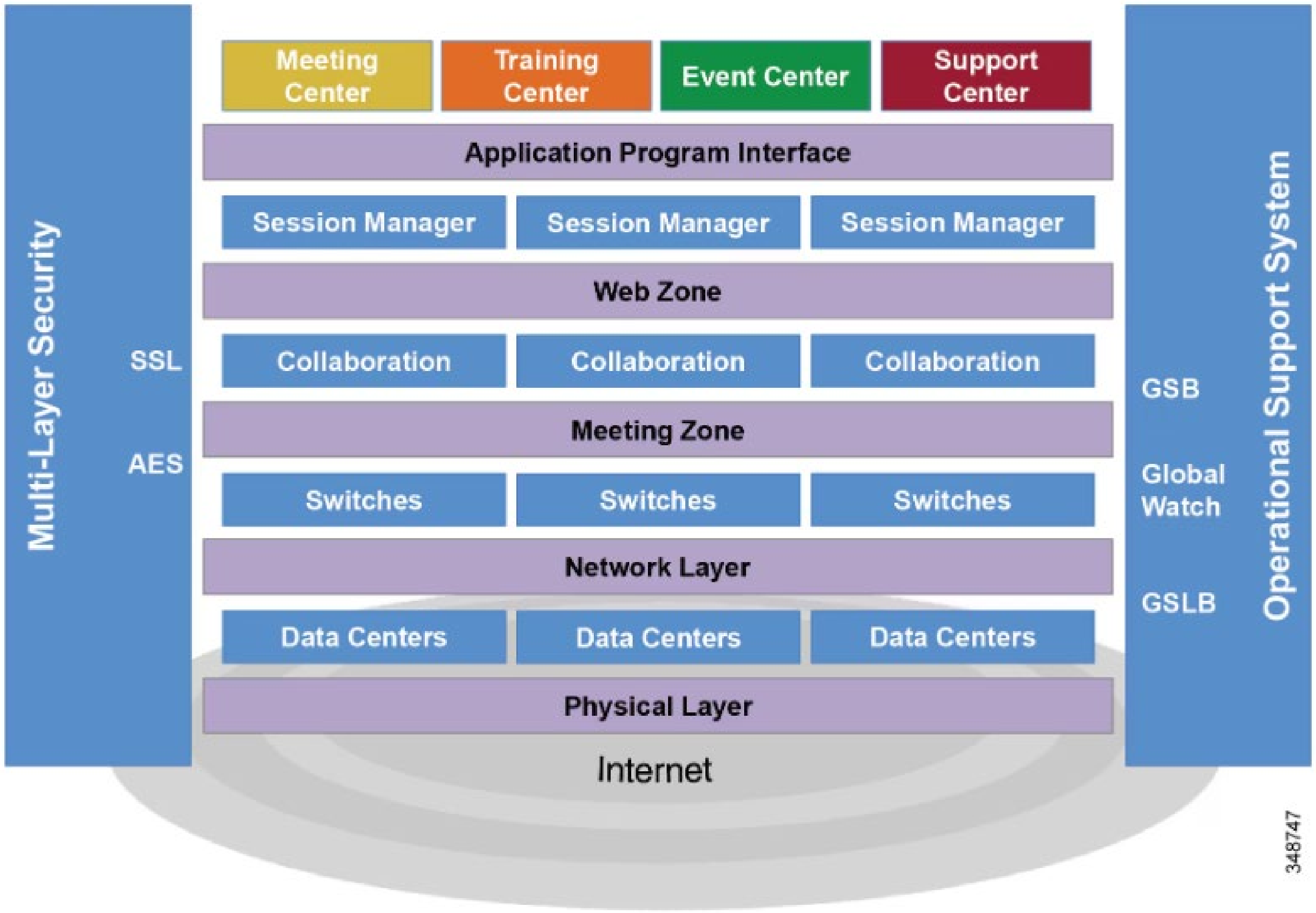
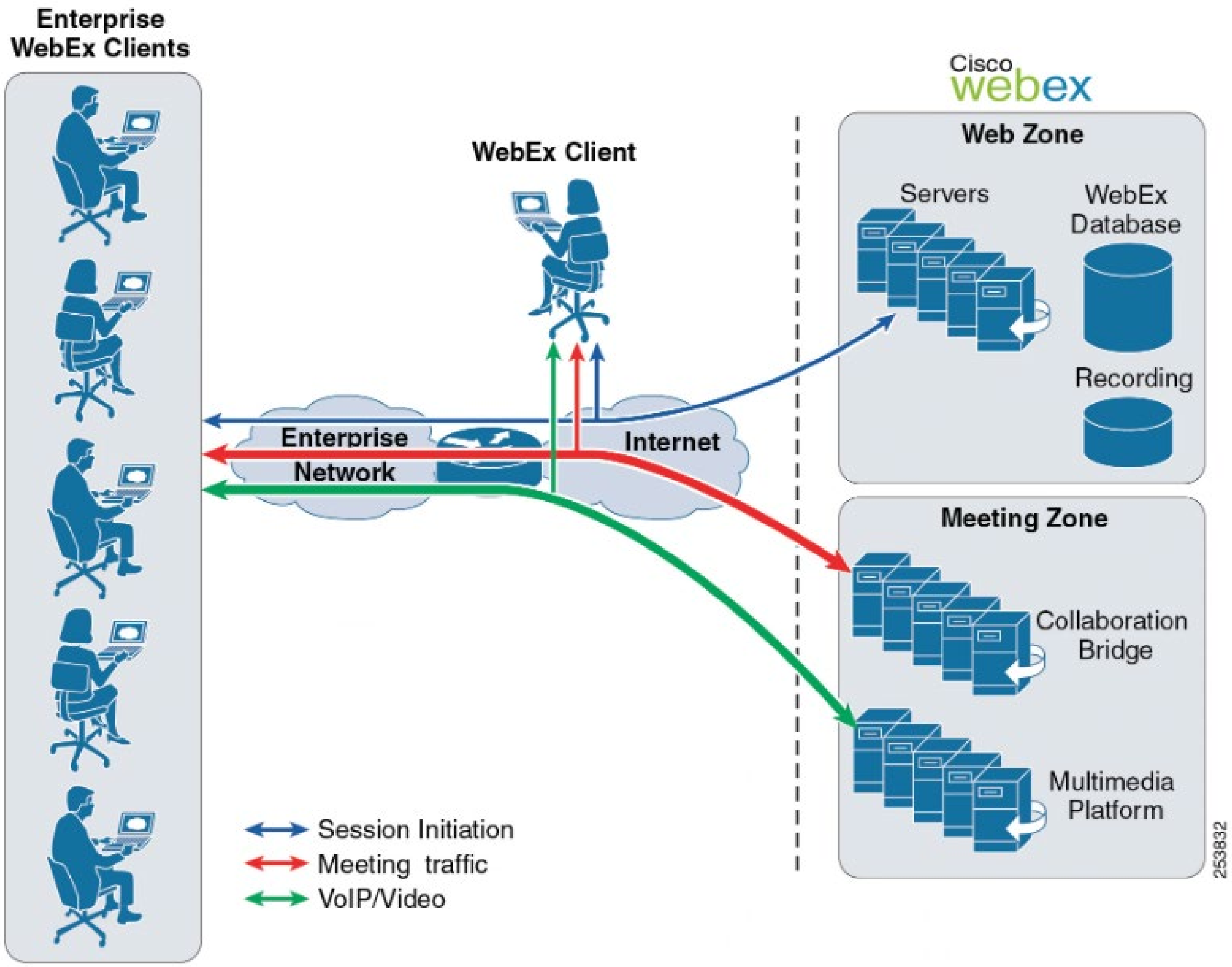
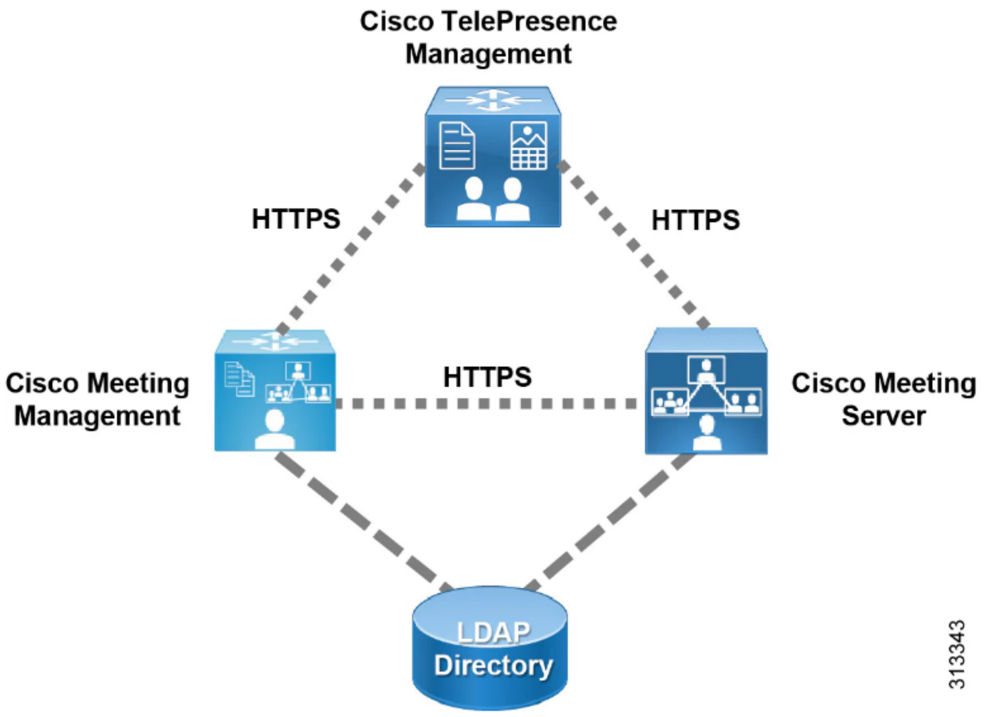
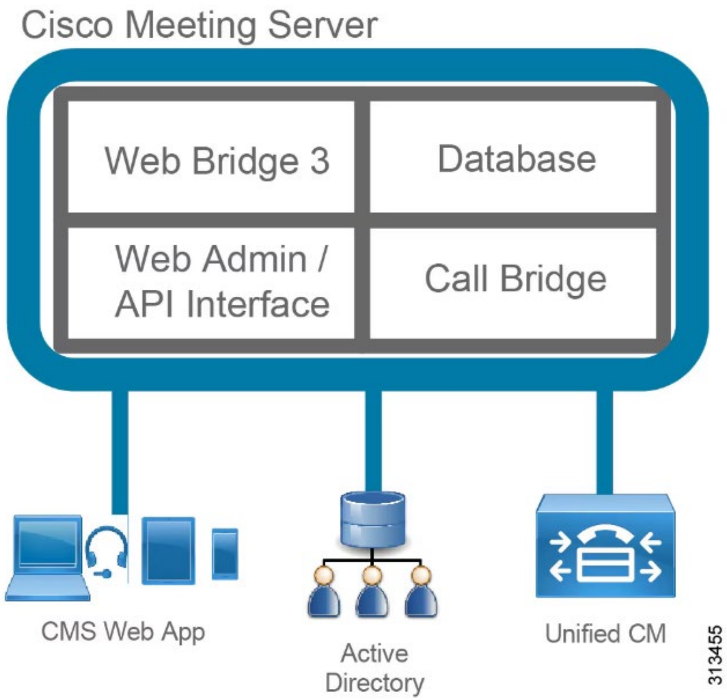


	Figure 11-15 WebEx Deployment 

	Cisco Rich Media Conferencing consists of the conferencing solutions described below. The details pertaining to each solution are described in each individual section that follows. • Cisco Unified CM Audio Conferencing This solution allows Unified CM to use its internal software component or external hardware digital signal processors (DSPs) as the resources to perform audio conferencing. • Cisco Meeting Server Cisco Meeting Server is an on-premises video conferencing solution. Each user has a personal Space that can be used to conduct meetings. Users can manage items such Space creation, adding members to a Space, and PIN creation from the Cisco Meeting App. • Cisco Collaboration Meeting Rooms Hybrid Cisco CMR Hybrid combines the on-premises video conference and the WebEx Meeting Center conference into a single meeting, which allows TelePresence and WebEx participants to join and share voice, video, and content. CMR Hybrid meetings can be either scheduled or non-scheduled. • Cisco WebEx Meeting Center Video Conferencing Cisco WebEx Meeting Center Video Conferencing (formerly Cisco Collaboration Meeting Rooms (CMR) Cloud) is an alternate conferencing deployment model that does not require any on-premises conferencing resources or management infrastructure. It supports both scheduled and non-scheduled meetings as well as TelePresence, audio, and WebEx participants in a single call, all hosted in the cloud. • Cisco WebEx Meetings Server Where cloud-based web and audio conferencing is not suitable, it is possible to use the on-premises WebEx Meetings Server solution. This product offers a standalone audio, video, and collaboration web conferencing platform. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab12/collab12/confernc.html.)

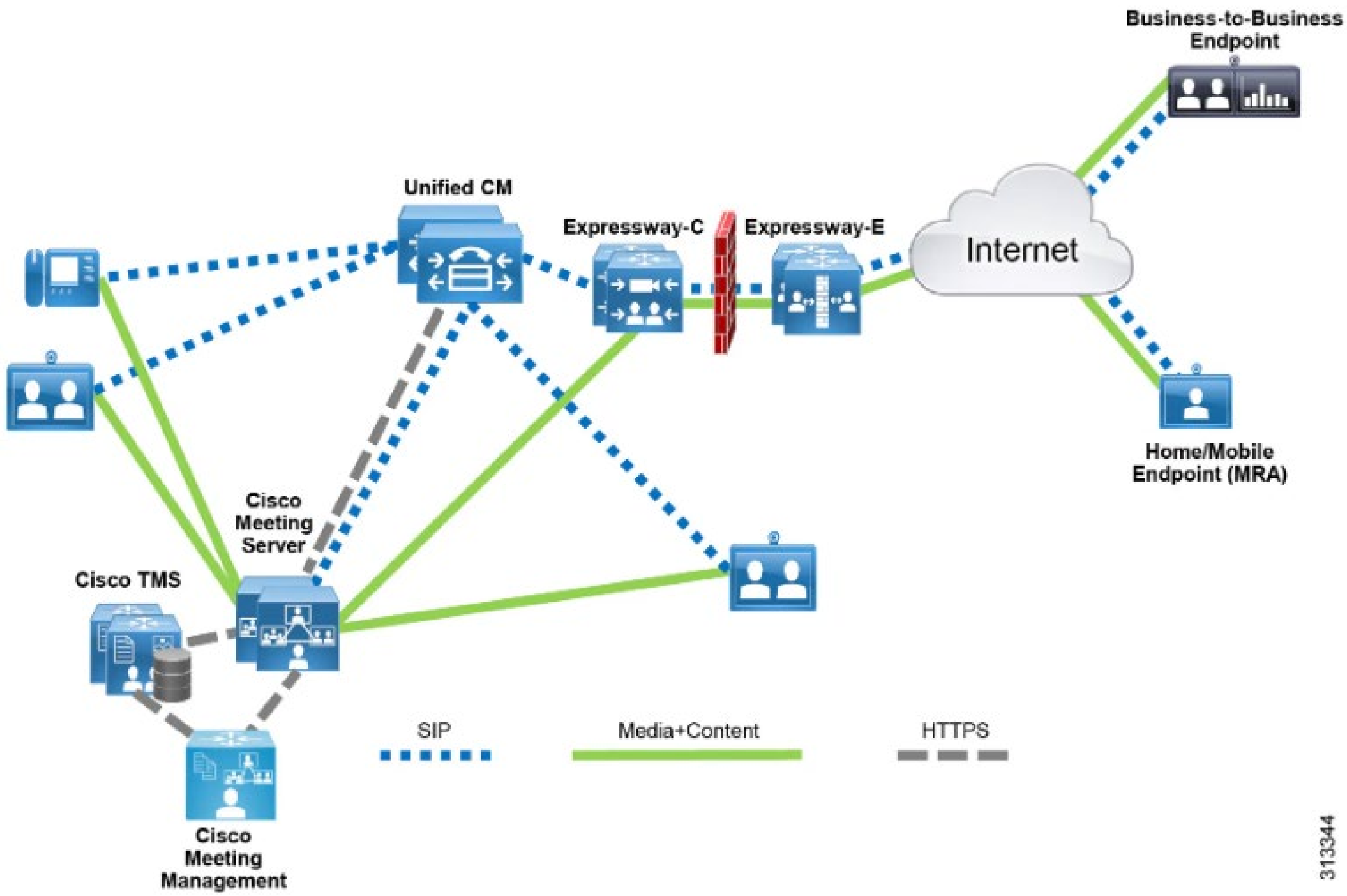
	## Network Traffic Planning Network traffic planning for Cisco WebEx Meeting Center Video Conferencing consists of the following elements: • WebEx Clients bandwidth The WebEx meeting client uses Scalable Video Coding (SVC) technology to send and receive video. It uses multi-layer frames to send video, and the receiving client automatically selects the best possible resolution to receive video that typically requires 1.2 to 3 Mbps of available bandwidth. For more information regarding network traffic planning for WebEx clients, refer to the *Cisco WebEx Network Bandwidth* white paper, available at https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meeting-center/white_paper_c11-691351.html • Bandwidth for video device from enterprise to WebEx Cloud For optimal SIP audio and video quality, Cisco recommends setting up the video bandwidth for at least 1.5 Mbps per device screen in the region associated with the endpoint registering with Cisco Unified CM. For example, if a triple-screen device registers with Unified CM, video bandwidth of 4.5 Mbps should be allocated in the associated region. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab12/collab12/confernc.html.)

	Figure 3-3 Cisco Meeting Management Architecture 

Cisco Meeting Server
Web Bridge 3
Database
Web Admin / API Interface
Call Bridge
CMS Web App
Active Directory
Unified CM
313455

Figure 3-5 Standard Deployment



(https://www.cisco.com/c/en/us/td/docs/solutions/CVD/Collaboration/enterprise/14/collbcvd/conferencing.html.)

How to deploy Cisco Meeting Server

Talk with a Cisco salesperson or partner to learn about deployment and choose the best options for you.

Select a platform

Cisco Meeting Server software has been optimized to run on our UCS-based Cisco Meeting Server 1000 and Cisco Meeting Server 2000

Choose a licensing option

Our multiparty option supports per-meeting licensing. Or you can purchase capacity units, as in a traditional license model.

Consider add-on features

Include recording ports, or consider Solution Plus partner Vbrick for recording/streaming distribution and Vyopta for assurance and analytics.

Download the software

Get the Cisco Meeting App for Macs and PCs on our site or from iTunes. Use the Apple Store for iOS devices.

Download now › Go to iTunes ›

Platforms

Cisco Meeting Server 1000

This Cisco UCS x86 server supports up to 120 simultaneous HD video conferencing calls.

Cisco Meeting Server 2000

This Cisco UCS x86 server supports up to 875 simultaneous HD video conferencing calls.

(https://www.cisco.com/c/en/us/products/conferencing/meeting-server/index.html.)

For example, the AV1 standard discloses a method for transmitting video signals (e.g., video bitstream). The AV1 standard is a video coding & decoding standard. It discloses that the AV1 standard is developed for video bitstream communication application like Internet related application, streaming, etc. It induces transmission of encoded video bitstream via Internet.

	AV1 Bitstream & Decoding Process Specification *Last modified: 2019-01-08 11:48 PT* *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 (Title Page)*, available at https://aomediacodec.github.io/av1-spec/av1-spec.pdf. Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.)

(http://dgql.org/~unlord/MHV2018.pdf.)

The Accused Instrumentalities implement the AV1 standard. For example:

	## Video Calls The typical video call includes two or three Real-Time Protocol (RTP) streams in each direction (that is, four or six streams). The call can include the following stream types: • Video (H.261, H.263, H.263+, H.264-SVC, X-H.264UC, H.264-AVC, H.265, AV1 and VT Camera wideband video codecs) • Far-End Camera Control (FECC) - Optional • Binary Floor Control Protocols (BFCP) **AV1 Codec Support** AV1 is a next-generation video codec developed by the Alliance for Open Media. The benefits of AV1 are: • Reduced bandwidth consumption and better visual quality by utilizing better compression efficiency compared to other video encodings • Enables video for users on very low bandwidth networks • Significant screen sharing efficiency improvements over other codecs Unified Communication Manager supports negotiation of AV1 codec to establish media if endpoints support the AV1 codec. When both endpoints support Multiple Codecs in Answer, Unified CM negotiates all the matching codecs including AV1 based on the preference order received. The endpoint will then use one of the codecs from the negotiated codec list for media streaming. In a low bandwidth environment, the AV1 codec is preferred by the endpoint over other codecs in the negotiated list. When both the endpoints involved in the call do not support the Multiple Codec in Answer, and the AV1 is the preferred codec over other codecs, Unified CM selects AV1 as the negotiated codec.

	SIP Video SIP video supports the following video calls by using the SIP Signaling Interface (SSI): • SIP to SIP • SIP to H.323 • SIP to SCCP • SIP intercluster trunk • H.323 trunk • Combination of SIP and H.323 trunk SIP video calls also provide media control functions for video conferencing. Unified Communications Manager video supports SIP on both SIP trunks and lines support video signaling. SIP supports the H.261, H.263, H.263+, H.264 (AVC), H.264 (SVC), X-H.264UC (Lync), and AV1 video codecs (it does not support the wideband video codec that the VTA uses). (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/14SU2/cucm_b_feature-configuration-guide-for-cisco14su2/cucm_m_video-telephony.html.) **AV1 Codec Support** Unified Communications Manager now supports negotiation and passthrough of AVI codec. The AV1 is a modern codec that provides better compression and hence can provide the same user experience as H.264 video codec at half the bandwidth. AV1 codec will be supported by Cisco Webex Desk Pro Endpoint, Webex Codec Pro, and Room Panorama systems. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/rel_notes/14_0_1/cucm_b_release-notes-for-cucm-imp-14_0_1/cucm_m_new-and-changed-features.html.)

	webex ahead Collaboration Workspaces Customer Experience Event Management Innovation & AI Sear **The AV1 video codec comes to Webex!** On Dec 15, 2020 - By Webex Team | 4 Min Read Thomas Davies & Sijia Chen – Webex is rolling out the AV1 video codec into production early next year. This will bring our media quality to the next level. As a founding member of AOM, Cisco is proud to introduce this advanced video technology into the real time communications market It's here! We have begun the process of rolling out the advanced AV1 video codec across Webex, taking video quality to the next level in the process, and replacing the aging H.264 standard.

	What do I need to use AV1 in Webex? Transmitting AV1 is supported when sharing screens or applications with "Optimize for motion and video" selected , and when the machine you are on has at least four cores. Receiving AV1 is supported for any machine with at least two cores. AV1 will automatically be used for sharing this type of screen content whenever all participants in a meeting support it, otherwise it will automatically revert to H.264.

How we are rolling out AV1

Adopting a brand-new video codec has an impact on every part of our Collaboration portfolio, so we are going step-by-step.

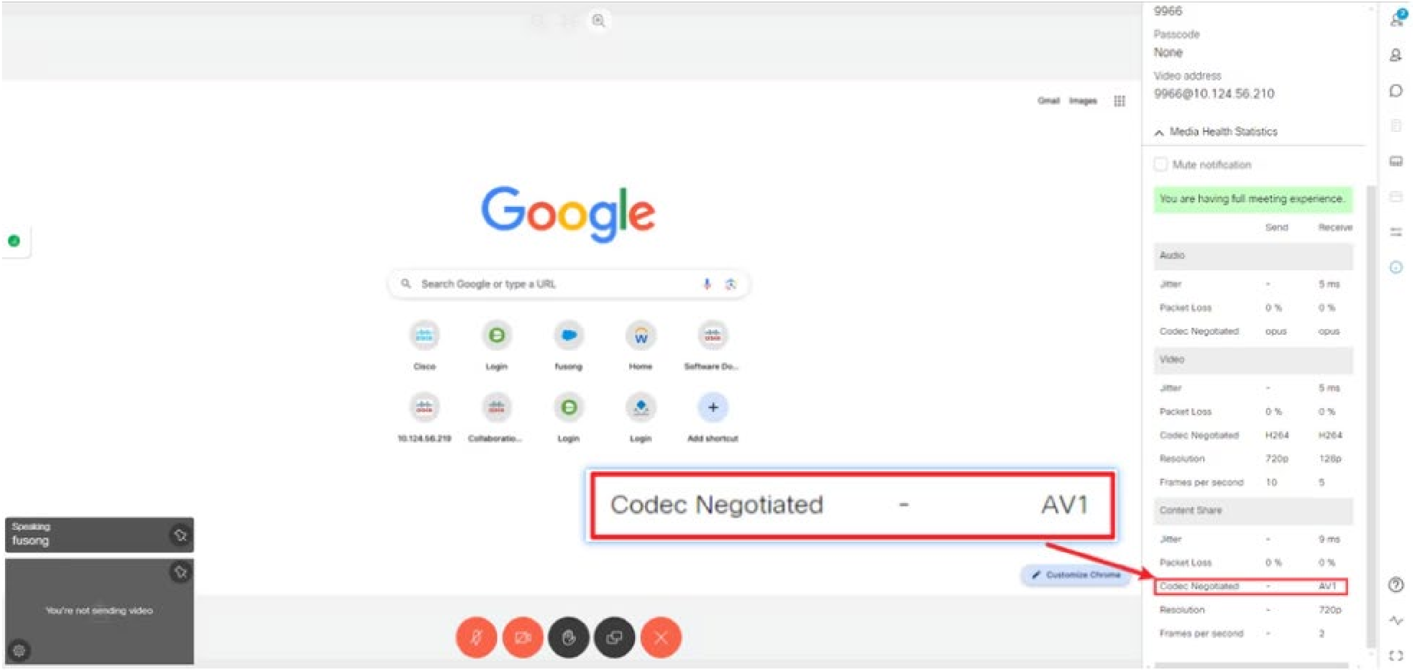
In future releases we will systematically expand where we deploy AV1. The immediate next steps are to support AV1 for other desktop share modes – either optimized for text and images, or automatically optimized. AV1 works just as well for these modes too, but we are being careful to change things gradually to make sure the user experience is perfect at each step.

Webex employs a fundamentally switched architecture, where video from each participant in a meeting is coded on their machine at different qualities and sent via a server to the other meeting participants. Initially, if some of those participants cannot support AV1 then we will automatically fall back to using H.264. Over time we will also remove these restrictions by applying ad hoc transcoding between AV1 and H.264 for those participants. This will also allow AV1 meetings to be recorded without reverting to H.264, for example.

Mobile devices will also rapidly gain hardware AV1 support, and then AV1 can be rolled out to mobile too. Although our solution is software-based and very fast on ARM as well as x86 processors, it is always better to make use of hardware codecs where possible on mobile to get the best battery life possible.

We'll also be seeking to reduce the restrictions we have placed on core count for AV1 as we continue to optimize. In fact, remarkably, our AV1 solution uses little more CPU than H.264. However, there are a huge range of different machines out there, and again we are moving gradually to safeguard user experience.

	(https://blog.webex.com/engineering/the-av1-video-codec-comes-to-webex.) "16 years after H.264, it's time for something new. Today, we demo'd an industry first: live, real-time AV1 encoding and transmission in a Webex meeting, with HD video & screen share!" — Anurag Dhingra, Cisco Webex CTO (https://medium.com/millicast/its-time-for-real-time-av1-video-encoding-withwebrtc-75a6aa64777c.) **AV1 Codec Support** Unified Communications Manager now supports negotiation and passthrough of AVI codec. The AV1 is a modern codec that provides better compression and hence can provide the same user experience as H.264 video codec at half the bandwidth. AV1 codec will be supported by Cisco Webex Desk Pro Endpoint, Webex Codec Pro, and Room Panorama systems. See the compatibility matrix for the compatible version of Webex Room devices, Cisco Expressway, and Cisco Meeting Server. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/rel_notes/14_0_1/cucm_b_release-notes-for-cucm-imp-14_0_1/cucm_m_new-and-changed-features.pdf.)

	2. The media health statistics of the content receiver show the content negotiate codec is AV1 on Chrome browser when receiving the content from CMS servers.  *Receiver content codec is AV1 on chrome* (https://www.cisco.com/c/en/us/support/docs/conferencing/meeting-server/221776-configure-av1-feature-on-cms.html.)
receiving a layered video data stream comprising a base layer and a set of enhancement layers;	The Accused Instrumentalities perform a method for transmitting video signals, comprising receiving a layered video data stream comprising a base layer and a set of enhancement layers. For example, the AV1 standard discloses receiving a layered video data stream (e.g., video bitstream) comprising a base layer (e.g., base layer) and a set of enhancement layers (e.g., enhancement layers).

	As shown below, the AV1 standard discloses an encoded video data bitstream using scalable video coding in a sequence of OBUs i.e., open bitstream unit. A metadata syntax of an OBU discloses scalability corresponding to the OBU. It discloses three types of scalabilities, Spatial scalability, Temporal scalability and Quality scalability. These scalabilities define a spatial layer having a corresponding spatial_id and a temporal layer having a corresponding temporal_id. Further, the AV1 standard discloses deriving a layered coded bitstream of base layer and enhancement layers using scalable video coding. It discloses a base layer having both spatial_id and temporal_id equal to zero and enhancement layers with either spatial_id or temporal_id equal to greater than zero values. **Bitstream** The sequence of bits generated by encoding a sequence of frames. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **5.8.1. General metadata OBU syntax**

metadata_obu( ) {	Type
metadata_type	leb128()
if (metadata_type == METADATA_TYPE_ITUT_T35)	
metadata_itut_t35( )	
else if (metadata_type == METADATA_TYPE_HDR_CLL)	
metadata_hdr_cll( )	
else if (metadata_type == METADATA_TYPE_HDR_MDCV)	
metadata_hdr_mdcv( )	
else if (metadata_type == METADATA_TYPE_SCALABILITY)	
metadata_scalability( )	
else if (metadata_type == METADATA_TYPE_TIMECODE)	
metadata_timecode( )	
}	

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 34 of 669) **5.8.5. Metadata scalability syntax** <table><tr><td>metadata_scalability( ) {</td><td>Type</td></tr><tr><td> scalability_mode_idc</td><td>f(8)</td></tr><tr><td> if (scalability_mode_idc == SCALABILITY_SS)</td><td></td></tr><tr><td> scalability_structure( )</td><td></td></tr><tr><td>}</td><td></td></tr></table> *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)*

5.8.6. Scalability structure syntax

scalability_structure( ) {	Type
spatial_layers_cnt_minus_1	f(2)
spatial_layer_dimensions_present_flag	f(1)
spatial_layer_description_present_flag	f(1)
temporal_group_description_present_flag	f(1)
scalability_structure_reserved_3bits	f(3)
if (spatial_layer_dimensions_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1 ; i++) {	
spatial_layer_max_width[i]	f(16)
spatial_layer_max_height[i]	f(16)
}	
}	
if (spatial_layer_description_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1; i++)	
spatial_layer_ref_id[i]	f(8)
}	
if (temporal_group_description_present_flag) {	
temporal_group_size	f(8)
for (i = 0; i < temporal_group_size; i++) {	
temporal_group_temporal_id[i]	f(3)
temporal_group_temporal_switching_up_point_flag[i]	f(1)
temporal_group_spatial_switching_up_point_flag[i]	f(1)
temporal_group_ref_cnt[i]	f(3)
for (j = 0; j < temporal_group_ref_cnt[i]; j++) {	
temporal_group_ref_pic_diff[i][j]	f(8)
}	

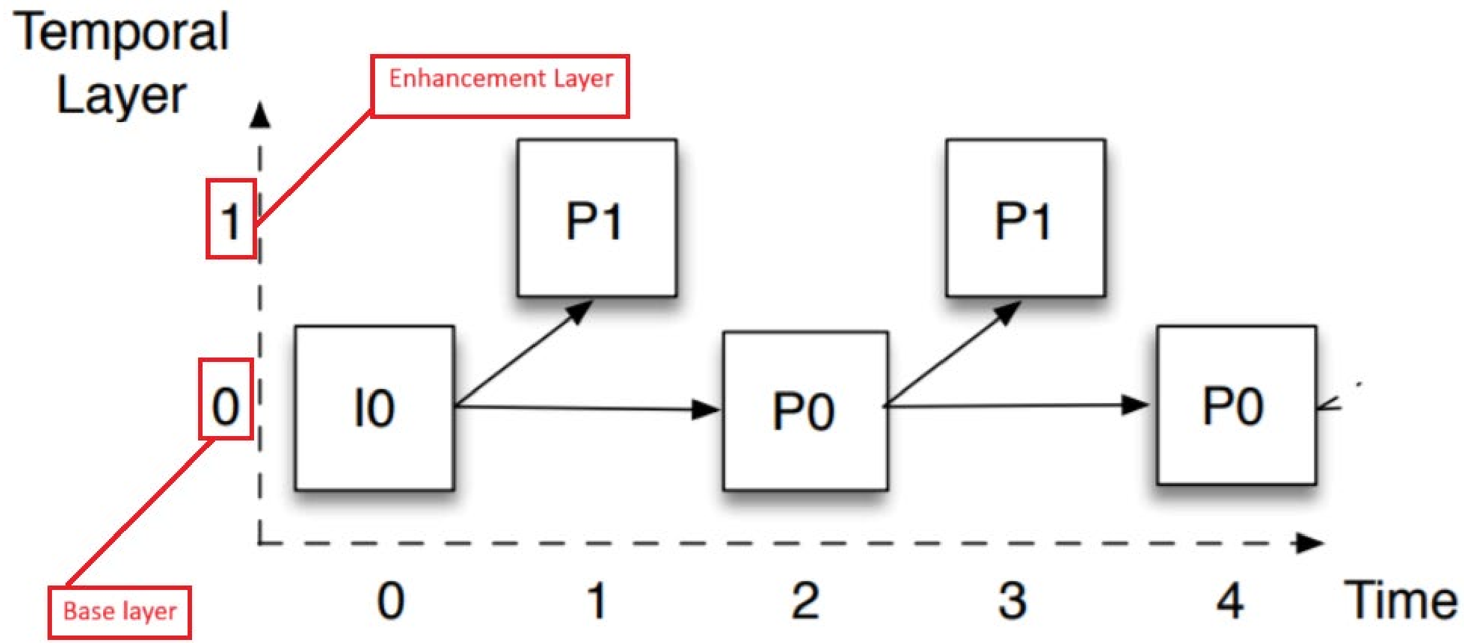
Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

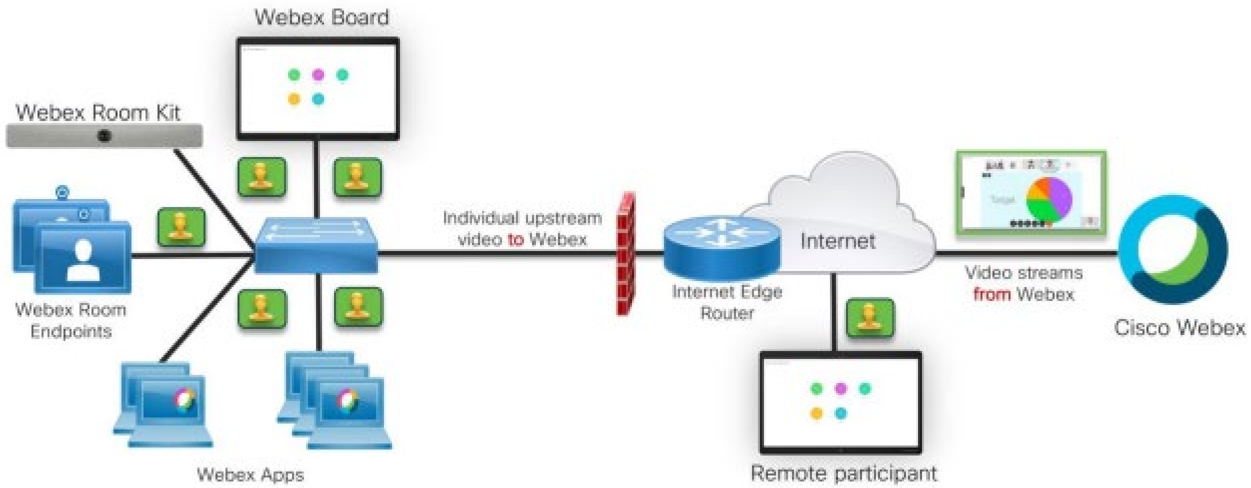
spatial_layer_ref_id[i] specifies the spatial_id value of the frame within the current temporal unit that the frame of layer i uses for reference. If no frame within the current temporal unit is used for reference the value must be equal to 255.

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669) Note that for a given picture, all frames follow the same inter-picture temporal dependency structure. However, the frame rate of each layer can be different from each other. The specified dependency structure in the scalability structure data must be for the highest frame rate layer. **temporal_group_temporal_id[i]** specifies the temporal_id value for the i-th picture in the temporal group. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669)* **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)*

	The AV1 codec can maintain up to eight reference frames, of which up to seven can be referenced by any new frame. AV1 also allows a frame to use another frame of a different spatial resolution as a reference frame. This allows internal resolution changes without requiring the use of key frames. These features together enable an AV1 encoder to implement various forms of coarse-grained scalability, including temporal, spatial, and quality scalability modes, as well as combinations of these, without the need for explicit scalable coding tools. Spatial and quality layers define different and possibly dependent representations of a single input frame. For a given spatial layer, temporal layers define different frame rates of video. Spatial layers allow a frame to be encoded at different spatial resolutions, whereas quality layers allow a frame to be encoded at the same spatial resolution but at different qualities (and thus with different amounts of coding error). AV1 supports quality layers as spatial layers without any resolution changes; hereinafter, the term “spatial layer” is used to represent both spatial and quality layers. This payload format specification provides for specific mechanisms through which such temporal and spatial scalability layers can be described and communicated. Temporal and spatial scalability layers are associated with non-negative integer IDs. The lowest layer of either type has an ID equal to 0. (https://aomediacodec.github.io/av1-rtp-spec/.) **Layer** A set of tile group OBUs with identical spatial_id and identical temporal_id values. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 3 of 669)* **Base layer** The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)*

	Enhancement layer A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)* As shown below, the AV1 standard discloses a base layer and enhancement layers for a coded video bitstream. **Note:** The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* **Note:** Examples are given for non-scalable cases, but the constraints also apply to each operating point of a scalable stream. For example, consider a 60fps spatial scalable stream with a base layer at 960x540, and an enhancement layer at 1920x1080. The operating point containing just the base layer may be labelled as level 3.0, while the operating point containing both the base and enhancement layer may be labelled as level 4.1. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 641 of 669)*

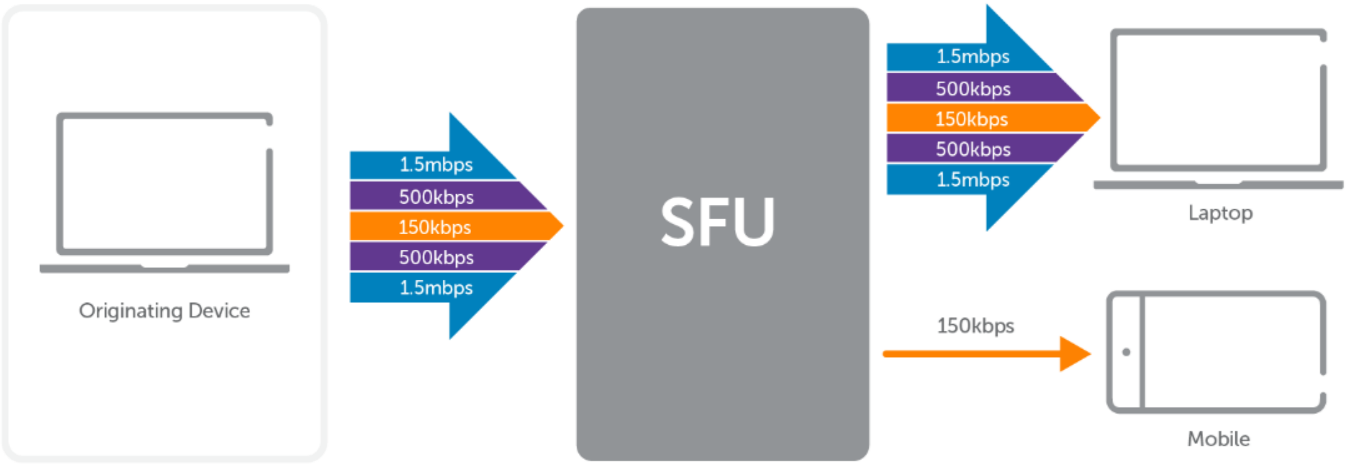
	 *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 126 of 669)*

	 ***Figure 11**: Video Streams in a Webex Meeting* (https://www.cisco.com/c/dam/td-xml/en_us/collaboration/cloud_cmr/pcia_2_0/reports/Troubleshooting_Audio_and_Video_Quality_Using_Webex_Control_Hub.pdf.)
identifying bandwidth-limited conditions of an internet protocol network between a video router and a plurality of video receivers;	The Accused Instrumentalities perform a method for transmitting video signals, comprising identifying bandwidth-limited conditions of an internet protocol network between a video router and a plurality of video receivers. For example, the AV1 standard discloses identifying bandwidth-limited conditions (e.g., network conditions, available bandwidth condition for a receiving device, etc.) of an internet protocol network (e.g., Internet, etc.) between a video router (e.g., a video data transmitter such as a video bitstream encoder, etc.) and a plurality of video receivers (e.g., video data receivers such as a video bitstream decoder, etc.).

	Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) For example, one of the applications of the AV1 standard discloses a selective forwarding unit or middlebox (SFU or SFM) which analyses bandwidth condition between a publisher/transmitter (e.g., video router) and a playback

	device (e.g., video receiver). According to the bandwidth condition, it selectively forwards layers to the playback device. **How Does SVC Work?** Video Router 1. **Publisher**: Transmits a single media stream (with multiple video layers) to the Selective Forwarding Unit (SFU). 2. **SFU**: Examines the stream and selectively encodes it into high and low bitrate layers. It then distributes these layers based on each playback device's network limitations and available bandwidth. 3. **Playback Devices**: Receive the appropriate layers, ensuring optimal quality for each participant. Video Receiver (https://www.linkedin.com/pulse/scalable-video-coding-svc-sudeep-kumar-nofdc.)

	The SFU then sends the adjusted stream along to the appropriate **playback device**. Depending on the available bandwidth and other limitations of the target devices, some could receive a lower quality stream while others get the whole onion (so to speak).  (https://www.wowza.com/blog/scalable-video-coding-for-webrtc.) **Selective Forwarding Middlebox (SFM)** A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant.

	(https://aomediacodec.github.io/av1-rtp-spec/.) As another example, the Accused Instrumentalities are capable of identifying bandwidth-limited conditions, such as latency, jitter, and/or packet loss between backend servers and a plurality of user clients:

Overview

In this document we will discuss bandwidth utilization. Bandwidth values used will be in payload bit rate which does not include packetization overhead and are covered in 3 categories, average, peak and maximum bit rate:

Average (**avg**) is the average over time for a meeting participant.

Peak (**peak**) is the typical peak bursts over the same time period for a meeting participant.

Maximum (**max**) is the maximum bit rate that the device is capable of either due to device limitations or device configuration.

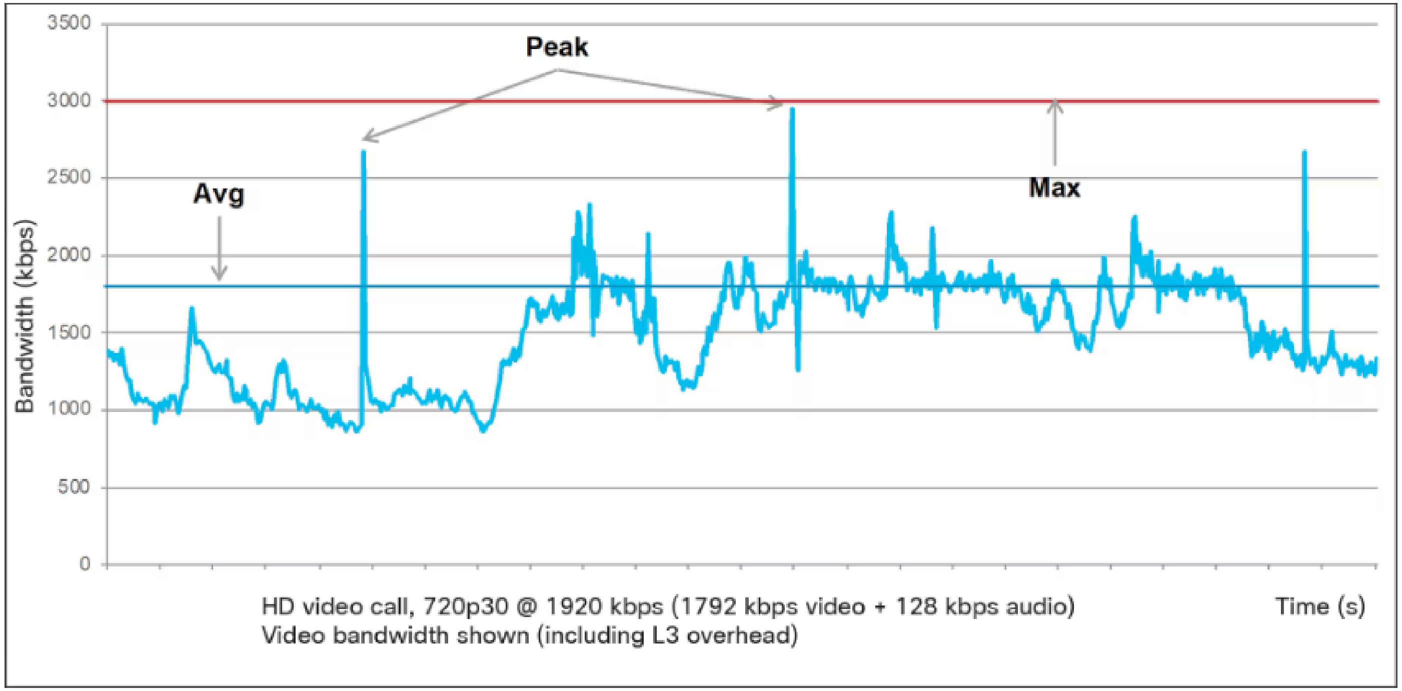
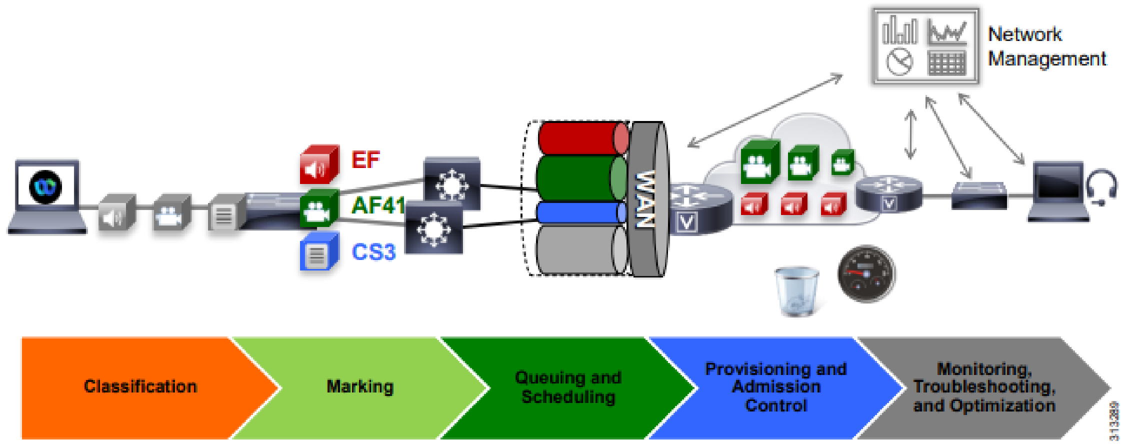


Figure 1.

Video Traffic: Bandwidth Usage High-definition Video Call

(https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings/white_paper_c11-691351.html.)

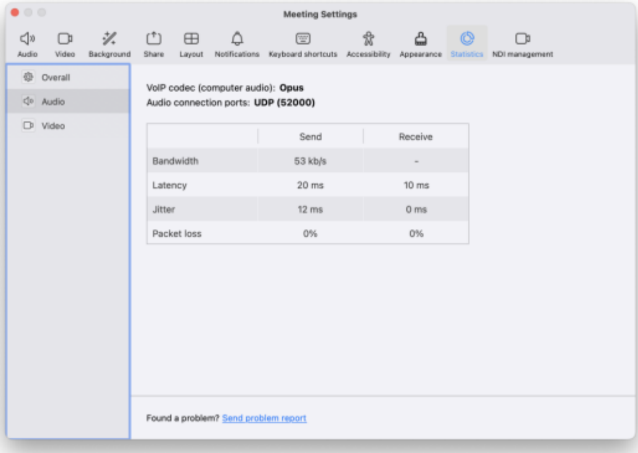
Figure 1 Architecture for Bandwidth Management

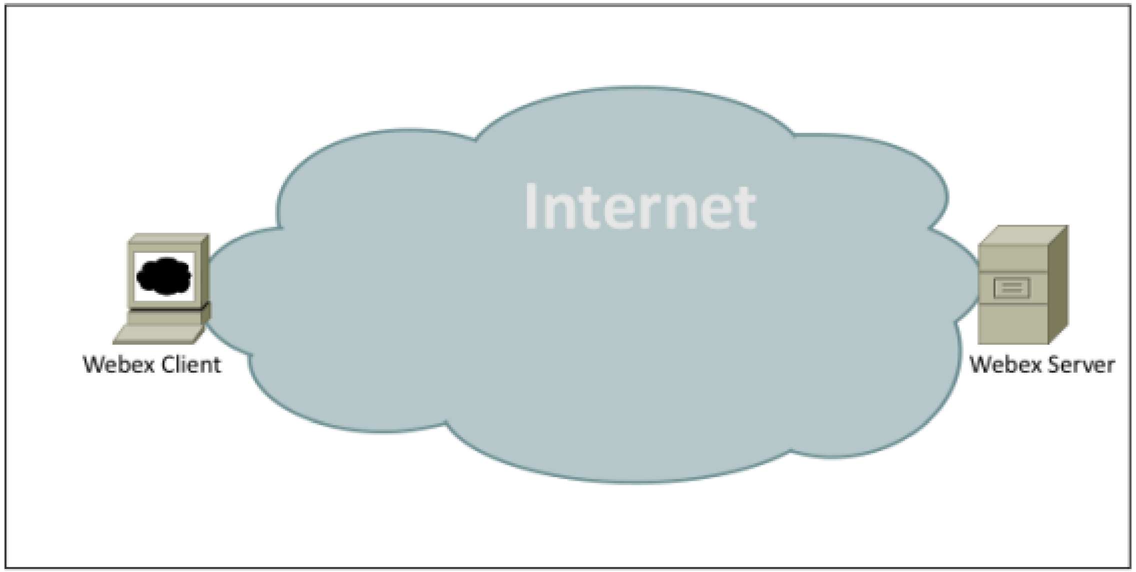


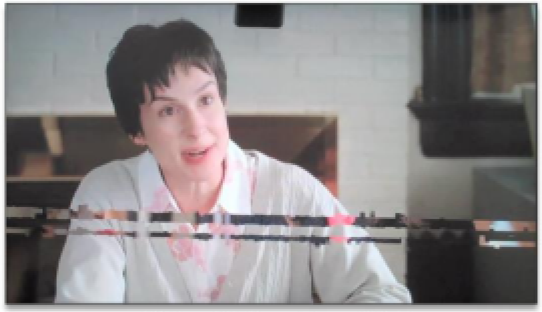
Recommended Deployment

Modify the existing on-premises QoS switch and WAN and Internet policies to include Webex Services identification, classification, and marking.

- Identify and classify media and signaling traffic from the Webex App and associated workloads as well as Webex Devices.
- Media and signaling marking recommendations:
  1. Mark all audio with Expedited Forwarding class EF (includes all audio of voice-only calls as well as audio for all types of video calls).
  2. Mark all Webex App video with an Assured Forwarding class of AF41 for a prioritized video class of service and AF42 for an opportunistic class of service. The marking of AF41 or AF42 will depend on the choice of whether to deploy opportunistic video during the on-premises deployment phase.
  3. Mark all call signaling with CS3. (All call signaling in HTTPS traffic will be marked based on the enterprise's current policy of traffic marking for HTTP/HTTPS – unless using NBAR which is covered in detail below)
- Configure QoS on all media originating and terminating applications such as the Video Mesh Nodes, Expressway and Cisco Unified Border Element.
- Update the WAN edge ingress re-marking policy.
- Update the WAN edge egress queuing and scheduling policy.

	Every Cisco video endpoint employs several smart media techniques to avoid network congestion, recover from packet loss, and optimize network resources. The following smart media techniques are just some of the techniques employed by Cisco Webex App and Devices: Media resilience techniques - Encoder pacing - Forward Error Correction (FEC) - Rate adaptation (https://www.cisco.com/c/dam/en/us/td/docs/solutions/CVD/Collaboration/AltDesigns/BWM-Wbx.pdf.) **Poor Audio / Video Quality** – Full-featured Meetings Help > Health Checker > Audio and Video Statistics... • Indicates TCP or UDP w/ Source Port • Latency / Packet Loss / Jitter  (https://www.ciscolive.com/c/dam/r/ciscolive/global-event/docs/2024/pdf/BRKCOL-3431.pdf.)

	There are three main factors that impact the quality of an audio or video call. These factors are **packet loss**, **latency**, and **jitter**. As shown in Figure 2, **packet loss is simply losing one or more packets within a stream of packets**. In this example, a Voice over IP (VoIP) packet is lost going from the Webex client to the Webex server. Loss can occur for a number of reasons in a network but most commonly it is caused by congestion or resource contention in the network or on the endpoints. This leads to routers, switches, or the endpoints themselves dropping or delaying packets. In other cases, packets encounter such a high delay that they are received too late to be played out. These packets will also be counted as lost.  ***Figure 2:*** *Packet Loss Example*

	***Note**: Control Hub Diagnostics highlights potential problems with packet loss and delay by coloring voice and video streams as Good, Fair, or Poor. It is important to understand that just because Control Hub Diagnostics flags part of a conversation as Fair or Poor, it does not necessarily mean that the user had a bad experience. Voice and video compensation mechanisms and algorithms in combination with other factors can mitigate the impact of loss and delay to the point where it is not noticeable by the user.* **Video Quality Artifacts** The first step in troubleshooting video quality is to clearly understand the symptom experienced by the end user. Getting a screenshot or a picture of the display screen usually gives more information than a verbal description of the problem. The type of video artifacts observed in a screenshot or picture can be used to determine the troubleshooting path. For example, in Figure 13, pictures (a) and (b) are artifacts that are caused when video streams experience packet loss. Just by seeing the picture, you can immediately start troubleshooting to identify the source of the packet loss.  a. Video with line stripe artifacts  b. Frozen video with block artifacts

Webex Control Hub makes it quite easy to find the device type and operating system version used by the end user to join a Webex meeting. All you need is the end user's email address and meeting time to find the right meeting and to get end user's device details. Figure 14 shows the list of meetings attended by a participant with the email address, rtpmsuser1@gmail.com.

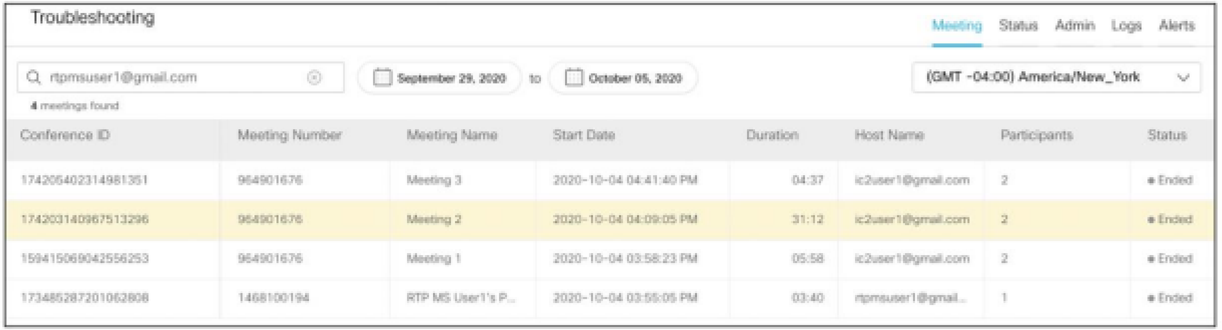


***Figure 14**: Participant Device Details*

(https://www.cisco.com/c/dam/td-xml/en_us/collaboration/cloud_cmr/pcia_2_0/reports/Troubleshooting_Audio_and_Video_Quality_Using_Webex_Control_Hub.pdf.)

	To support more than four video streams across a distribution link, it is recommended that the bandwidth of the link be set to greater than 2Mbps. Use the API or the Web Admin Interface to set the bandwidth. If using the API, PUT a value for the `peerLinkBitRate` parameter to the API object `/system/configuration/cluster`; the value will be the maximum media bit rate to use on distribution links between Call Bridges in the cluster. Alternatively, using the Web Admin Interface, go to **Configuration > Cluster > Call Bridge identity** and enter the **Peer link bit rate**. (https://www.cisco.com/c/dam/en/us/td/docs/conferencing/ciscoMeetingServer/Deployment_Guide/Version-3-9/Cisco-Meeting-Server-3-9-Scalable-and-Resilient-Deployment.pdf.)
forwarding the base layer to at least two of the plurality of video receivers via the internet protocol network; and selectively forwarding one or more of the set of enhancement layers, but fewer than all of the set of enhancement layers, to at least two of the plurality of video receivers through the internet protocol network based upon the identified bandwidth-limited conditions;	The Accused Instrumentalities perform a method for transmitting video signals, comprising forwarding the base layer to at least two of the plurality of video receivers via the internet protocol network; and selectively forwarding one or more of the set of enhancement layers, but fewer than all of the set of enhancement layers, to at least two of the plurality of video receivers through the internet protocol network based upon the identified bandwidth-limited conditions. For example, the AV1 standard discloses forwarding (e.g., selective forwarding using SFU – selective forwarding unit or SFM – selective forwarding middlebox) the base layer (e.g., base layer) to at least two of the plurality of video receivers (e.g., video data receiver such as a video bitstream decoder, etc.) via the internet protocol network (e.g., Internet, etc.) and selectively forwarding (e.g., selective forwarding using SFU – selective forwarding unit or SFM – selective forwarding middlebox) one or more of the set of enhancement layers (e.g., enhancement layers), but fewer than all of the set of enhancement layers (e.g., enhancement layers), to at least two of the plurality of video receivers (e.g., video data receiver such as a video bitstream decoder, etc.) through the internet protocol network (e.g., Internet, etc.) based upon the identified bandwidth-limited conditions (e.g., network condition, available bandwidth condition for a receiving device, etc.). The AV1 standard is a video coding & decoding standard. It discloses that the AV1 standard is developed for video bitstream communication application like Internet related application, streaming, etc. It induces transmission of encoded video bitstream via Internet. It also discloses scaling according to varying bandwidth condition.

	Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

AV1 Features

ROYALTY-FREE

Interoperable and open

UBIQUITOUS

Scales to any modern device at any bandwith

FLEXIBLE

For use in both commercial and non-commercial content, including user-generated content

30% BETTER COMPRESSION *

Uses less data while delivering 4k UHD video and beyond when compared to alternatives

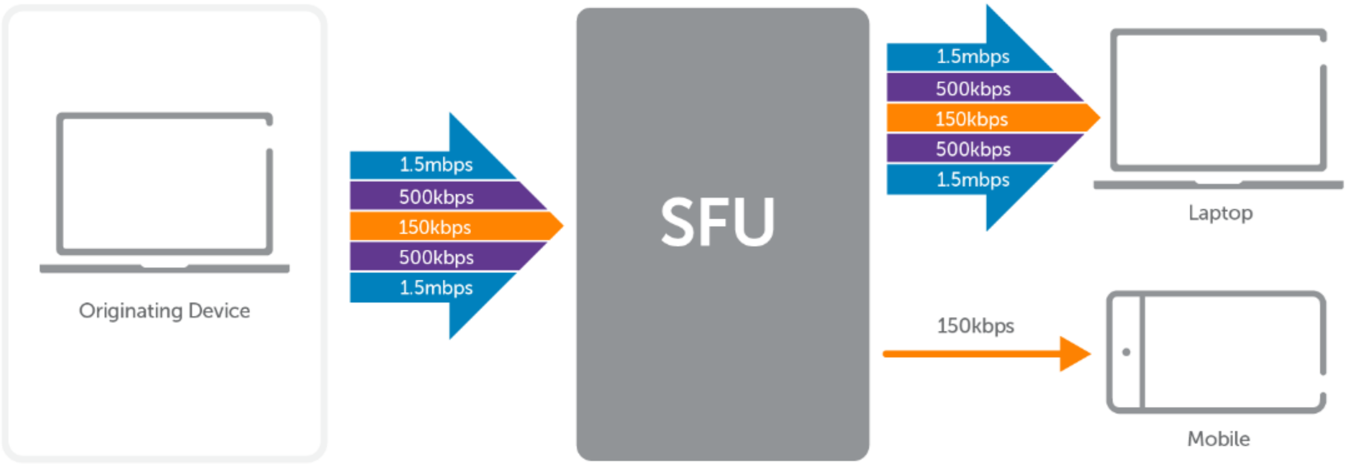
OPTIMIZED

Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services

(https://aomedia.org/av1-features/.)

For example, one of the applications of the AV1 standard discloses a selective forwarding unit or middlebox (SFU or SFM) which analyses bandwidth condition between a publisher/transmitter (e.g., video router) and a playback

	device (e.g., video receiver). According to the bandwidth condition, it selectively forwards layers to the playback device. As shown below, it forwards only base layer to devices which are having low bandwidth available and other devices depending on their bandwidth conditions could receive different number of enhancement layers. **How Does SVC Work?** Video Router 1. **Publisher**: Transmits a single media stream (with multiple video layers) to the Selective Forwarding Unit (SFU). 2. **SFU**: Examines the stream and selectively encodes it into high and low bitrate layers. It then distributes these layers based on each playback device's network limitations and available bandwidth. 3. **Playback Devices**: Receive the appropriate layers, ensuring optimal quality for each participant. Video Receiver (https://www.linkedin.com/pulse/scalable-video-coding-svc-sudeep-kumar-nofdc.)

	The SFU then sends the adjusted stream along to the appropriate **playback device**. Depending on the available bandwidth and other limitations of the target devices, some could receive a lower quality stream while others get the whole onion (so to speak).  (https://www.wowza.com/blog/scalable-video-coding-for-webrtc.)

	Selective Forwarding Middlebox (SFM) A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant. **Dependency Descriptor RTP Header Extension** **A.1 Introduction** This appendix describes the Dependency Descriptor (DD) RTP Header extension. The DD is used for conveying dependency information about individual video frames in a scalable video stream. The DD includes provisions for both temporal and spatial scalability. In the DD, the smallest unit for which dependencies are described is an RTP frame. An RTP frame contains one complete coded video frame and may also contain additional information (e.g., metadata). Each RTP frame is identified by a frame_number. When spatial scalability is used, there may be multiple RTP frames produced for the same time instant. Further, this specification allows for the transmission of an RTP frame over multiple packets. RTP frame aggregation is explicitly disallowed. Hereinafter, RTP frame will be referred to as frame. The DD uses the concept of Decode targets, each of which represents a subset of a scalable video stream necessary to decode the stream at a certain temporal and spatial fidelity. A frame may be associated with several Decode targets. This concept is used to facilitate selective forwarding, as is done by a Selective Forwarding Middlebox (SFM). Typically an SFM would select one Decode target for each endpoint, and forward all frames associated with that target.

6 MANE and SFM Behavior

If a packet contains an OBU with an OBU extension header then the entire packet is considered associated with the layer identified by the temporal_id and spatial_id combination that are indicated in the extension header. If a packet does not contain any OBU with an OBU extension header, then it is considered to be associated with all operating points.

The general function of a MANE or SFM is to selectively forward packets to receivers. To make forwarding decisions a MANE may inspect the media payload, so that it may need to be able to parse the AV1 bitstream and if so, cannot function when end-to-end encryption is enabled. An SFM does not parse the AV1 bitstream and therefore needs to obtain the information relevant to selective forwarding by other means, such as the Dependency Descriptor described in Appendix A. In addition to enabling bitstream-independent forwarding and support for end-to-end encryption, the Dependency Descriptor also enables forwarding where the metadata OBU provided in the AV1 bitstream is not sufficient to express the structure of the stream.

6.1.1 Example

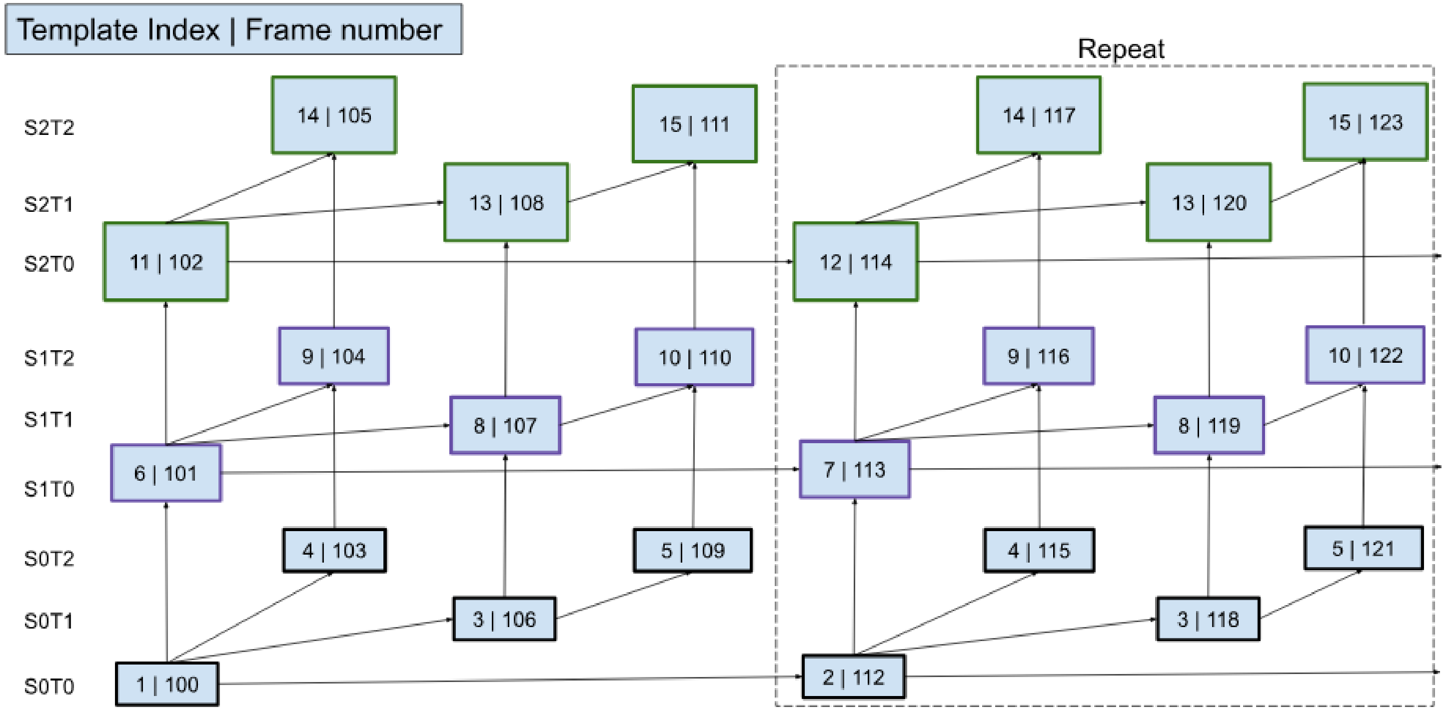
In this example, it is desired to send three simulcast encodings, each containing three temporal layers. When sending all encodings on a single SSRC, scalability mode 'S3T3' would be indicated within the scalability metadata OBU, and the Dependency Descriptor describes the dependency structure of all encodings.

When sending each simulcast encoding on a distinct SSRC, the scalability mode 'L1T3' would be indicated within the scalability metadata OBU of each bitstream, and the Dependency Descriptor in each stream describes only the dependency structure for that individual encoding. A distinct spatial_id (e.g. 0, 1, 2) could be used for each stream (if a single AV1 encoder is used to produce the three simulcast encodings), but if distinct AV1 encoders are used, the spatial_id values may not be distinct.

	<table><tr><th></th><th>Indication</th><th>Description</th><th>SFM behavior</th></tr><tr><td>DT0</td><td>Not present</td><td>F5 is not associated with DT0.</td><td>Do not forward F5 to a DT0 client.</td></tr><tr><td>DT1</td><td>Discardable</td><td>No frame in DT1 will reference F5.</td><td>Should forward F5 to a DT1 client, but may discard it, e.g., when bandwidth is low.</td></tr><tr><td>DT2</td><td>Switch</td><td>If it is possible to decode F5, then all future frames in DT2 are also decodable.</td><td>Forward F5 to the DT2 client. In addition, the SFM can rely on the fact that if the F5 frame is decodable for a particular client then that client may be switched to DT2.</td></tr><tr><td>DT3</td><td>Required</td><td>Future frames in DT3 reference F5.</td><td>Forward F5 to a DT3 client. No additional decisions can be made.</td></tr></table>

A.10.2.2 L3T3 Full SVC

This example uses three Chains. Chain 0 includes frames with spatial ID equal to 0 and temporal ID equal to 0. Chain 1 includes frames with spatial ID equal to 0 or 1 and temporal ID equal to 0. Chain 2 includes all frames with temporal ID equal to 0.



Idx	S	T	Fdiffs	Chains			Decode Targets								
				0	1	2	HD30 fps	HD15 fps	HD7.5 fps	VGA30 fps	VGA15 fps	VGA7.5fps	QVGA30 fps	QVGA15 fps	QVGA7.5 fps
1	0	0		0	0	0	S	S	S	S	S	S	S	S	S
2	0	0	12	12	11	10	R	R	R	R	R	R	S	S	S
3	0	1	6	6	5	4	R	R	-	R	R	-	S	D	-
4	0	2	3	3	2	1	R	-	-	R	-	-	D	-	-
5	0	2	3	9	8	7	R	-	-	R	-	-	D	-	-
6	1	0	1	1	1	1	S	S	S	S	S	S	-	-	-
7	1	0	12,1	1	1	1	R	R	R	S	S	S	-	-	-
8	1	1	6,1	7	6	5	R	R	-	S	D	-	-	-	-
9	1	2	3,1	4	3	2	R	-	-	D	-	-	-	-	-
10	1	2	3,1	10	9	8	R	-	-	D	-	-	-	-	-
11	2	0	1	2	1	1	S	S	S	-	-	-	-	-	-
12	2	0	12,1	2	1	1	S	S	S	-	-	-	-	-	-
13	2	1	6,1	8	7	6	S	D	-	-	-	-	-	-	-
14	2	2	3,1	5	4	3	D	-	-	-	-	-	-	-	-
15	2	2	3,1	11	10	9	D	-	-	-	-	-	-	-	-
decode_target_protected_by							2	2	2	1	1	1	0	0	0

(https://aomediacodec.github.io/av1-rtp-spec/.)

	Base layer The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **Enhancement layer** A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)* 6.2.3. OBU extension header semantics **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)*

	Note: The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). Coded video data of a temporal level with temporal_id T and spatial level with spatial_id S are only allowed to reference previously coded video data of temporal_id T' and spatial_id S', where T' <= T and S' <= S. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* If a coded video sequence contains at least one enhancement layer (OBUs with spatial_id greater than 0 or temporal_id greater than 0) then all frame headers and tile group OBUs associated with base (spatial_id equals 0 and temporal_id equals 0) and enhancement layer (spatial_id greater than 0 or temporal_id greater than 0) data must include the OBU extension header. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 205 of 669)*

Table 4. Webex Meetings Bandwidth per Resolution Table

Layer	Bandwidth Range
90p active thumbnail (each)	~60-100 kb/s
180p main video	125-200 kb/s
360p main video	470-640 kb/s
720p main video	900k-1.5 mb/s
Content sharing (sharpness, 1080p/5)	120k – 1.3 mb/s
Content sharing (motion, 720p/30)	900k – 2.5 mb/s

Webex Meetings Desktop App Bandwidth Controls

Webex administrators have 2 key controls to help control bandwidth as used by clients that connect to Webex meetings should they choose to. Namely, you can cap the meeting layouts at either 360p as the max available resolution, or to enable 720p layers. Whether your site is administered on Webex Control Hub or Webex Site Administrator, the following controls are available in Configuration > Common Site Settings > Options:

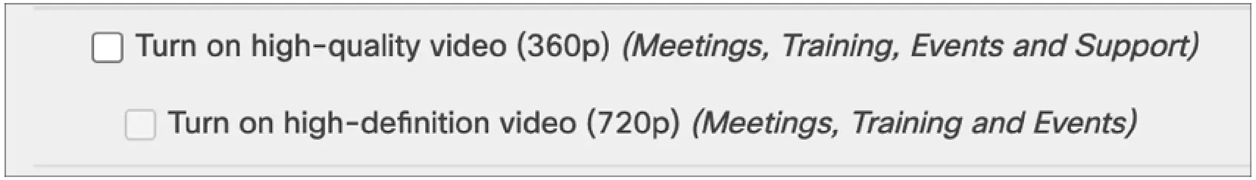


Figure 5.

Webex Meetings Desktop App Bandwidth Controls

Webex Media Improvements

The following are media improvements that have occurred in releases from 40.7 – 40.10

Improved Video Experience over WIFI (Defer video down-speeding): WIFI environments present a specific set of challenges and through testing we noticed the algorithms were too quick to drop resolution/bandwidth in certain WIFI environments. In 40.7 the algorithms have been updated to 'defer the down-speeding" of video for brief WIFI specific network dropouts. This resulted in improved user experience over WIFI. See figure5 for an illustration of this. When latency spikes for a specific duration the improved algorithms can detect that this is a WIFI specific event and not thus delay the down-speeding of the video resolution until that duration (n) is exceeded in which case would infer a different network event increasing latency.

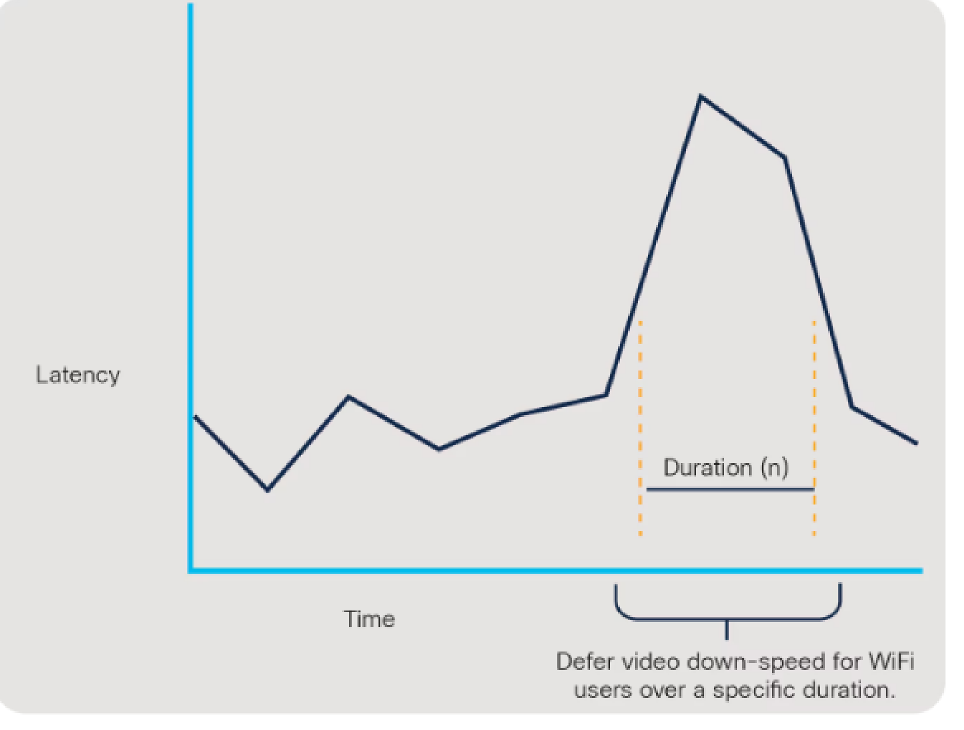


Figure 6.

Deferred Video Down-speeding

Video Super Scaling is a technique used for users on sub-optimal networks. We provide 720p like experiences to users who are experiencing network impairments resulting in 360p connections to our servers for their meeting. Our updated scaling algorithm provides enhanced video quality when this degradation occurs, making network impairment downgrades less quality impacting. If a 360p image is simply scaled geometrically to a 720p resolution, the quality has not improved, but our technique involves migrating video post-processing tasks to the client for some video flows versus doing it in the cloud, resulting in being able to keep more precious bits of information to reconstruct the video at a given bandwidth.

In a meeting as shown in figure 8 and 9, the bandwidth requirements are as follows. Assuming the receiver is not bandwidth restricted, Webex app will adapt to available network conditions. Table 8 outlines the bandwidth utilization. Figure 13 illustrates the Active Speaker layout which is the default meeting layout for Webex Teams App.

Table 8. Cisco Webex Teams Bandwidth Utilization

Meeting Scenario	Main Video+Audio (no content sharing)	Main Video+Audio (content sharing)
Webex Teams 1:1 Call	1.5 mb/s peak, 1.3 mb/s avg	1.5 mb/s peak, 1.3 mb/s avg
Webex Teams Meeting (fig 13 layout), 6 participants	2.3 mb/s peak, 1.7 mb/s avg	2.9/ mb/s peak, 2.0 mb/s avg
Webex Teams Meeting (fig 14 layout), 6 participants	2.5 mb/s peak, 1.8 mb/s avg	3.5 mb/s peak, 2.3 mb/s avg

(https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings/white_paper_c11-691351.html.)

wherein the layered video data stream is transmitted according to an internet protocol; and	The Accused Instrumentalities perform a method for transmitting video signals, wherein the layered video data stream is transmitted according to an internet protocol. For example, the AV1 standard discloses the method such that the layered video data stream (e.g., scalable video bitstream, etc.) is transmitted according to an internet protocol (e.g., Internet, etc.). Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.) **Selective Forwarding Middlebox (SFM)** A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant. (https://aomediacodec.github.io/av1-rtp-spec/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.)
wherein each layer of the layered video data stream comprises data packets, each	The Accused Instrumentalities perform a method for transmitting video signals, wherein each layer of the layered video data stream comprises data packets, each of which is encoded with a sequence number and a layer identifier, and wherein the layer identifier for each data packet is based upon a layer to which the packet belongs.

of which is encoded with a sequence number and a layer identifier, and wherein the layer identifier for each data packet is based upon a layer to which the packet belongs.	For example, the AV1 standard discloses the method such that each layer of the layered video data stream comprises data packets (e.g., IP packets of video bitstream data units) each of which is encoded with a sequence number (e.g., identification number value of IP packet, etc.) and a layer identifier (e.g., a layer identifier such as base layer, enhancement layer, etc.) and wherein the layer identifier (e.g., the layer identifier such as base layer, enhancement layer, etc.) for each data packet (e.g., IP packets of video bitstream data units) is based upon a layer (e.g., layer such as base layer, enhancement layer, etc.) to which the packet belongs. As shown below, the AV1 standard discloses an encoded video data bitstream using scalable video coding in a sequence of OBUs i.e., open bitstream unit. The OBU data units are transmitted in packetized format over Internet. These packets are governed by IP protocol and communicated as an IP packet. An IP packet comprises a header part and a payload/data part. The header part of the IP packet comprises an identification field which denotes a sequence number of IP packets i.e., data packets, transmitted.

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) **Bitstream** The sequence of bits generated by encoding a sequence of frames. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)*

OBU

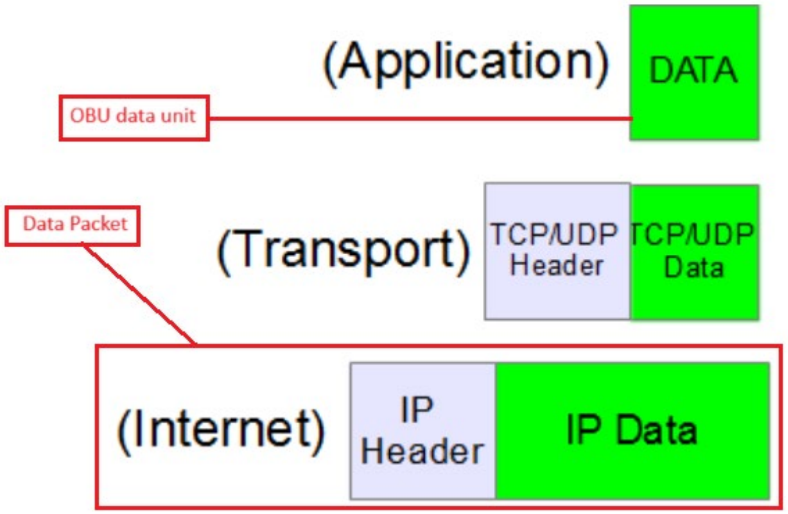
All structures are packetized in “Open Bitstream Units” or OBUs. Each OBU has a header, which provides identifying information for the contained data (payload).

Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page – 4 of 669)

Open Bitstream Unit (OBU)

The smallest bitstream data framing unit in AV1. All AV1 bitstream structures are packetized in OBUs.

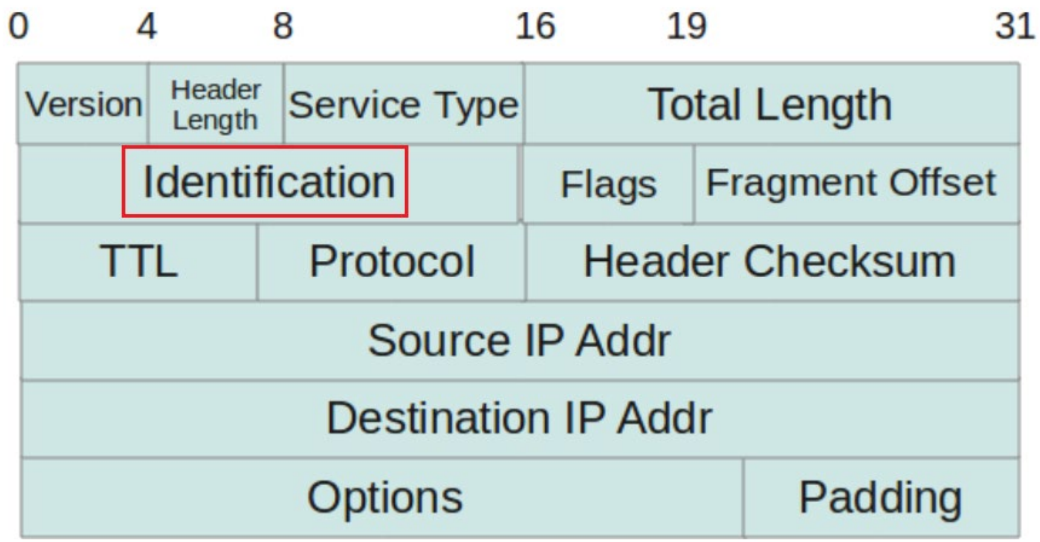
(https://aomediacodec.github.io/av1-rtp-spec/#3-media-format-description.)



(https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

	These are a set of standard rules that allows different types of computers to communicate with each other. The IP protocol ensures that each computer that is connected to the Internet is having a specific serial number called the IP address. TCP specifies how data is exchanged over the internet and how it should be broken into IP packets. It also makes sure that the packets have information about the source of the message data, the destination of the message data, the sequence in which the message data should be re-assembled, and checks if the message has been sent correctly to the specific destination. The TCP is also known as a connection-oriented protocol. (https://www.geeksforgeeks.org/types-of-internet-protocols/.)

IP Header



(https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

- ***Identification(16 bits)***: This field is used for uniquely identifying the IP datagrams. This value is incremented every-time an IP datagram is sent from source to the destination. This field comes in handy while reassembly of fragmented IP data grams.

(https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

Further, the OBU data unit comprises a metadata syntax which discloses scalability corresponding to the OBU. It discloses three types of scalabilities i.e., Spatial scalability, Temporal scalability and Quality scalability. These

scalabilities define a spatial layer having a corresponding spatial_id and a temporal layer having a corresponding temporal_id.

Further, the AV1 standard discloses deriving a layered coded bitstream of base layer and enhancement layers using scalable video coding. It discloses a base layer having both spatial_id and temporal_id equal to zero and enhancement layers with at least one of spatial_id or temporal_id values greater than zero.

Bitstream

The sequence of bits generated by encoding a sequence of frames.

Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)

5.8.1. General metadata OBU syntax

metadata_obu( ) {	Type
metadata_type	leb128()
if (metadata_type == METADATA_TYPE_ITUT_T35)	
metadata_itut_t35( )	
else if (metadata_type == METADATA_TYPE_HDR_CLL)	
metadata_hdr_cll( )	
else if (metadata_type == METADATA_TYPE_HDR_MDCV)	
metadata_hdr_mdcv( )	
else if (metadata_type == METADATA_TYPE_SCALABILITY)	
metadata_scalability( )	
else if (metadata_type == METADATA_TYPE_TIMECODE)	
metadata_timecode( )	
}	

Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 34 of 669)

	5.8.5. Metadata scalability syntax

metadata_scalability( ) {	Type
scalability_mode_idc	f(8)
if (scalability_mode_idc == SCALABILITY_SS)	
scalability_structure( )	
}	

Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

5.8.6. Scalability structure syntax

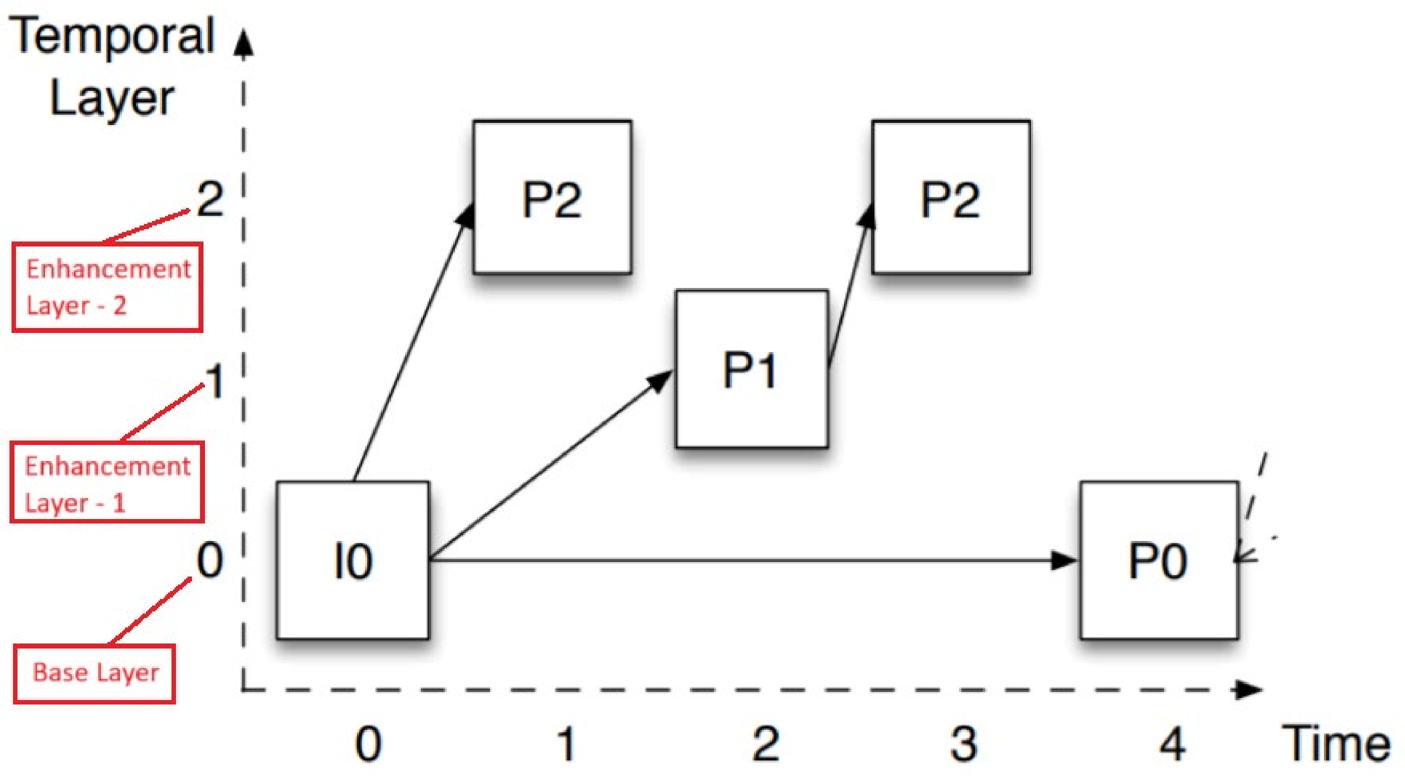
scalability_structure( ) {	Type
spatial_layers_cnt_minus_1	f(2)
spatial_layer_dimensions_present_flag	f(1)
spatial_layer_description_present_flag	f(1)
temporal_group_description_present_flag	f(1)
scalability_structure_reserved_3bits	f(3)
if (spatial_layer_dimensions_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1 ; i++) {	
spatial_layer_max_width[i]	f(16)
spatial_layer_max_height[i]	f(16)
}	
}	
if (spatial_layer_description_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1; i++)	
spatial_layer_ref_id[i]	f(8)
}	
if (temporal_group_description_present_flag) {	
temporal_group_size	f(8)
for (i = 0; i < temporal_group_size; i++) {	
temporal_group_temporal_id[i]	f(3)
temporal_group_temporal_switching_up_point_flag[i]	f(1)
temporal_group_spatial_switching_up_point_flag[i]	f(1)
temporal_group_ref_cnt[i]	f(3)
for (j = 0; j < temporal_group_ref_cnt[i]; j++) {	
temporal_group_ref_pic_diff[i][j]	f(8)
}	

Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

spatial_layer_ref_id[i] specifies the spatial_id value of the frame within the current temporal unit that the frame of layer i uses for reference. If no frame within the current temporal unit is used for reference the value must be equal to 255.

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669) Note that for a given picture, all frames follow the same inter-picture temporal dependency structure. However, the frame rate of each layer can be different from each other. The specified dependency structure in the scalability structure data must be for the highest frame rate layer. **temporal_group_temporal_id[i]** specifies the temporal_id value for the i-th picture in the temporal group. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669)* **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)* **Base layer** The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **Enhancement layer** A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)*

	Note: The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). Coded video data of a temporal level with temporal_id T and spatial level with spatial_id S are only allowed to reference previously coded video data of temporal_id T' and spatial_id S', where T' <= T and S' <= S. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* **Note:** Examples are given for non-scalable cases, but the constraints also apply to each operating point of a scalable stream. For example, consider a 60fps spatial scalable stream with a base layer at 960x540, and an enhancement layer at 1920x1080. The operating point containing just the base layer may be labelled as level 3.0, while the operating point containing both the base and enhancement layer may be labelled as level 4.1. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 641 of 669)*



Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 126 of 669)

Dependency Descriptor RTP Header Extension

A.1 Introduction

This appendix describes the Dependency Descriptor (DD) RTP Header extension. The DD is used for conveying dependency information about individual video frames in a scalable video stream. The DD includes provisions for both temporal and spatial scalability.

In the DD, the smallest unit for which dependencies are described is an RTP frame. An RTP frame contains one complete coded video frame and may also contain additional information (e.g., metadata). Each RTP frame is identified by a frame_number. When spatial scalability is used, there may be multiple RTP frames produced for the same time instant. Further, this specification allows for the transmission of an RTP frame over multiple packets. RTP frame aggregation is explicitly disallowed. Hereinafter, RTP frame will be referred to as frame.

The DD uses the concept of Decode targets, each of which represents a subset of a scalable video stream necessary to decode the stream at a certain temporal and spatial fidelity. A frame may be associated with several Decode targets. This concept is used to facilitate selective forwarding, as is done by a Selective Forwarding Middlebox (SFM). Typically an SFM would select one Decode target for each endpoint, and forward all frames associated with that target.

A.8 Dependency Descriptor Format

To facilitate the work of selectively forwarding portions of a scalable video bitstream, as is done by an SFM, for each packet, the following information is made available (even though not all elements are present in every packet).

- spatial ID
- temporal ID
- DTIs
- frame_number of the current frame
- frame_number of each of the Referred frames
- frame_number of last frame in each Chain

6 MANE and SFM Behavior

If a packet contains an OBU with an OBU extension header then the entire packet is considered associated with the layer identified by the temporal_id and spatial_id combination that are indicated in the extension header. If a packet does not contain any OBU with an OBU extension header, then it is considered to be associated with all operating points.

The general function of a MANE or SFM is to selectively forward packets to receivers. To make forwarding decisions a MANE may inspect the media payload, so that it may need to be able to parse the AV1 bitstream and if so, cannot function when end-to-end encryption is enabled. An SFM does not parse the AV1 bitstream and therefore needs to obtain the information relevant to selective forwarding by other means, such as the Dependency Descriptor described in Appendix A. In addition to enabling bitstream-independent forwarding and support for end-to-end encryption, the Dependency Descriptor also enables forwarding where the metadata OBU provided in the AV1 bitstream is not sufficient to express the structure of the stream.

6.1.1 Example

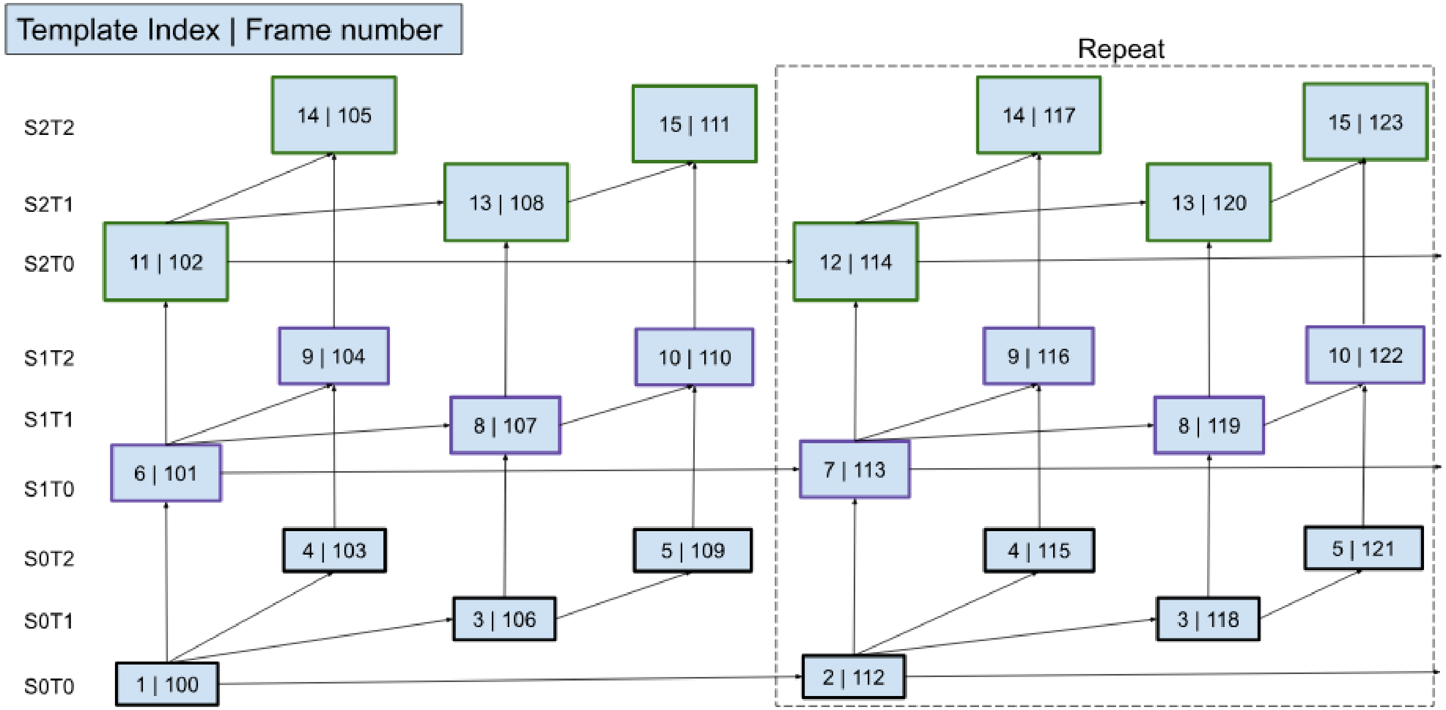
In this example, it is desired to send three simulcast encodings, each containing three temporal layers. When sending all encodings on a single SSRC, scalability mode 'S3T3' would be indicated within the scalability metadata OBU, and the Dependency Descriptor describes the dependency structure of all encodings.

When sending each simulcast encoding on a distinct SSRC, the scalability mode 'L1T3' would be indicated within the scalability metadata OBU of each bitstream, and the Dependency Descriptor in each stream describes only the dependency structure for that individual encoding. A distinct spatial_id (e.g. 0, 1, 2) could be used for each stream (if a single AV1 encoder is used to produce the three simulcast encodings), but if distinct AV1 encoders are used, the spatial_id values may not be distinct.

	Indication	Description	SFM behavior
DT0	Not present	F5 is not associated with DT0.	Do not forward F5 to a DT0 client.
DT1	Discardable	No frame in DT1 will reference F5.	Should forward F5 to a DT1 client, but may discard it, e.g., when bandwidth is low.
DT2	Switch	If it is possible to decode F5, then all future frames in DT2 are also decodable.	Forward F5 to the DT2 client. In addition, the SFM can rely on the fact that if the F5 frame is decodable for a particular client then that client may be switched to DT2.
DT3	Required	Future frames in DT3 reference F5.	Forward F5 to a DT3 client. No additional decisions can be made.

A.10.2.2 L3T3 Full SVC

This example uses three Chains. Chain 0 includes frames with spatial ID equal to 0 and temporal ID equal to 0. Chain 1 includes frames with spatial ID equal to 0 or 1 and temporal ID equal to 0. Chain 2 includes all frames with temporal ID equal to 0.



	(table below)

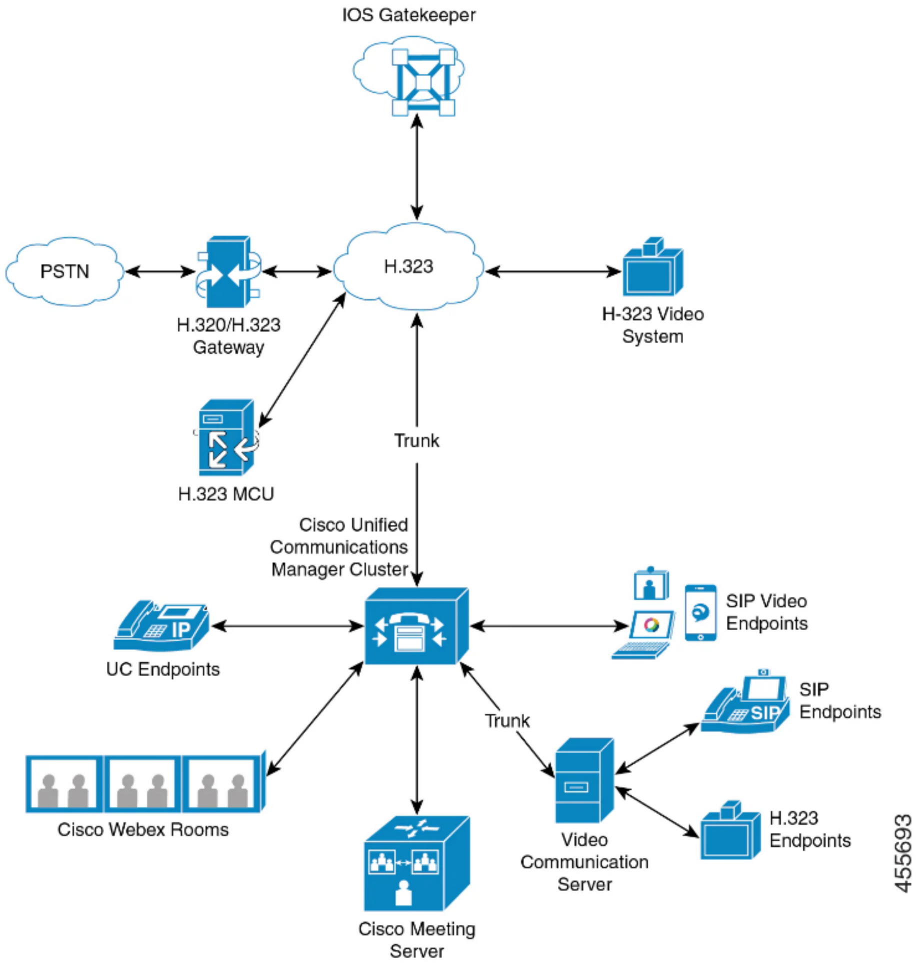
Idx	S	T	Fdiffs	Chains			Decode Targets								
				0	1	2	HD30 fps	HD15 fps	HD7.5 fps	VGA30 fps	VGA15 fps	VGA7.5fps	QVGA30 fps	QVGA15 fps	QVGA7.5 fps
1	0	0		0	0	0	S	S	S	S	S	S	S	S	S
2	0	0	12	12	11	10	R	R	R	R	R	R	S	S	S
3	0	1	6	6	5	4	R	R	-	R	R	-	S	D	-
4	0	2	3	3	2	1	R	-	-	R	-	-	D	-	-
5	0	2	3	9	8	7	R	-	-	R	-	-	D	-	-
6	1	0	1	1	1	1	S	S	S	S	S	S	-	-	-
7	1	0	12,1	1	1	1	R	R	R	S	S	S	-	-	-
8	1	1	6,1	7	6	5	R	R	-	S	D	-	-	-	-
9	1	2	3,1	4	3	2	R	-	-	D	-	-	-	-	-
10	1	2	3,1	10	9	8	R	-	-	D	-	-	-	-	-
11	2	0	1	2	1	1	S	S	S	-	-	-	-	-	-
12	2	0	12,1	2	1	1	S	S	S	-	-	-	-	-	-
13	2	1	6,1	8	7	6	S	D	-	-	-	-	-	-	-
14	2	2	3,1	5	4	3	D	-	-	-	-	-	-	-	-
15	2	2	3,1	11	10	9	D	-	-	-	-	-	-	-	-
decode_target_protected_by							2	2	2	1	1	1	0	0	0

(https://aomediacodec.github.io/av1-rtp-spec/.)

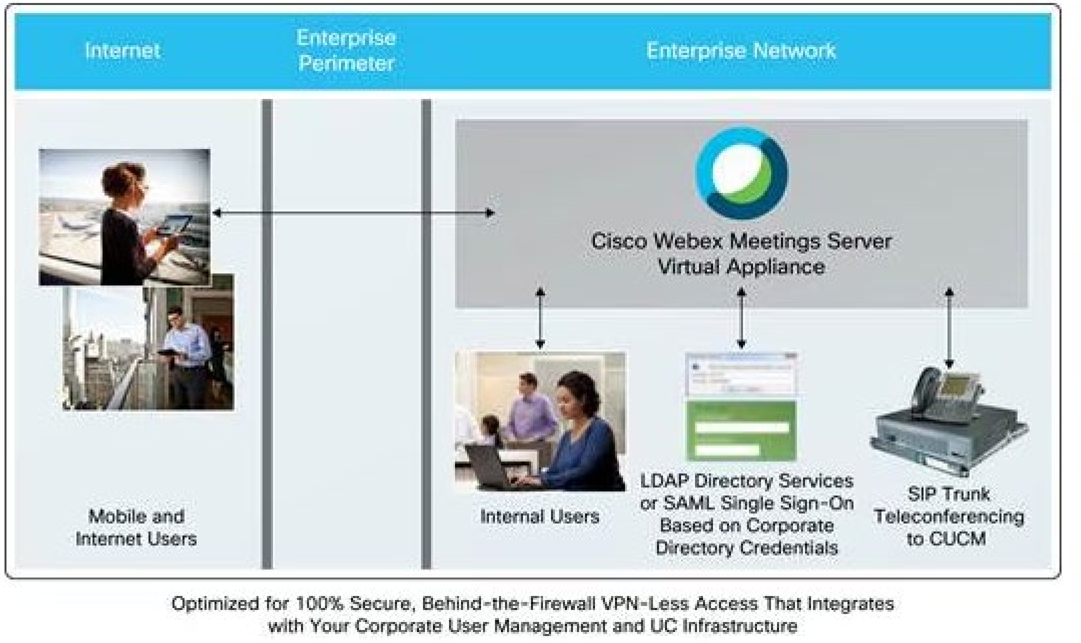
11. A scalable video coding router, comprising:	The Accused Instrumentalities include a scalable video coding router. For example:

Video Network

The following illustration provides an example of a video network that uses a single Unified Communications Manager cluster. In a successful video network, any endpoint can call any other endpoint. Video availability only exists if both endpoints are video-enabled. Video capabilities extend across trunks.



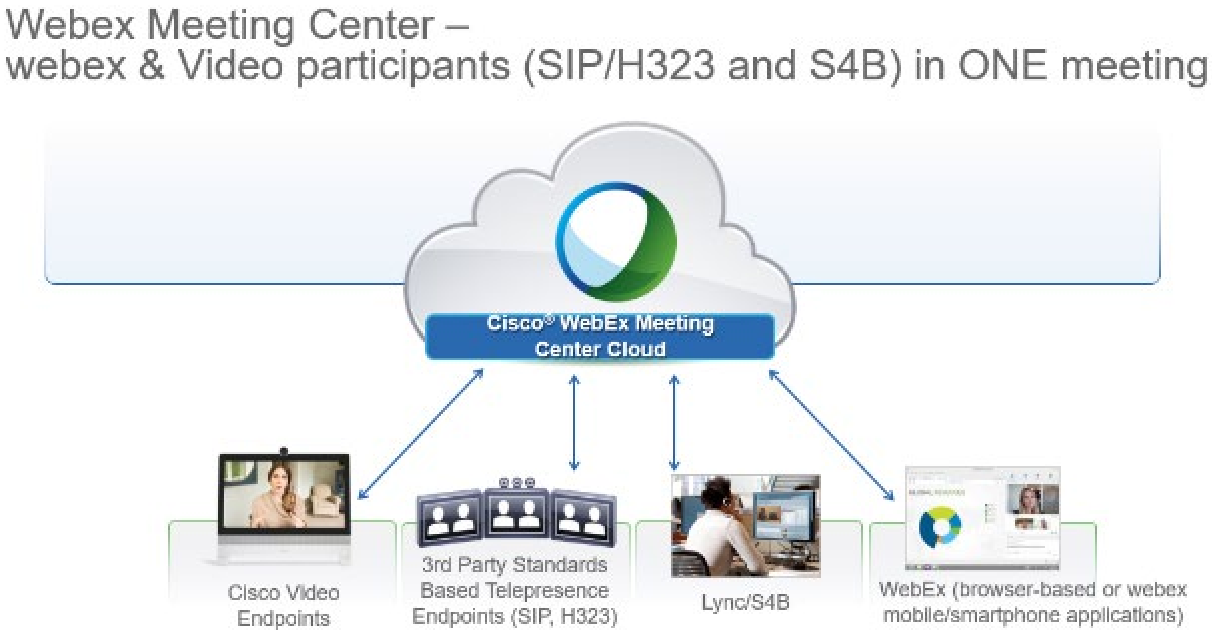
	The Cisco video conference portfolio comprises the following video bridges: • Cisco TelePresence MCU series • Webex Meeting Server (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/14SU2/cucm_b_feature-configuration-guide-for-cisco14su2/cucm_m_video-telephony.html.) **Product Overview** Cisco Webex Meetings Server is a virtualized, software-based solution that runs on Cisco Unified Computing System™ (Cisco UCS®) servers and VMware. It uses virtual appliance technology for rapid turn-up of services to end users. With Cisco Webex Meetings Server, there are two options for enabling mobile users to more securely access Cisco Webex conferences without going through a VPN. The first option is to deploy reverse proxy (or edge servers) in the enterprise perimeter (or DMZ). The second option, shown in Figure 1, is to deploy the reverse proxy servers behind your internal firewall, thus eliminating all DMZ components and related information security concerns.

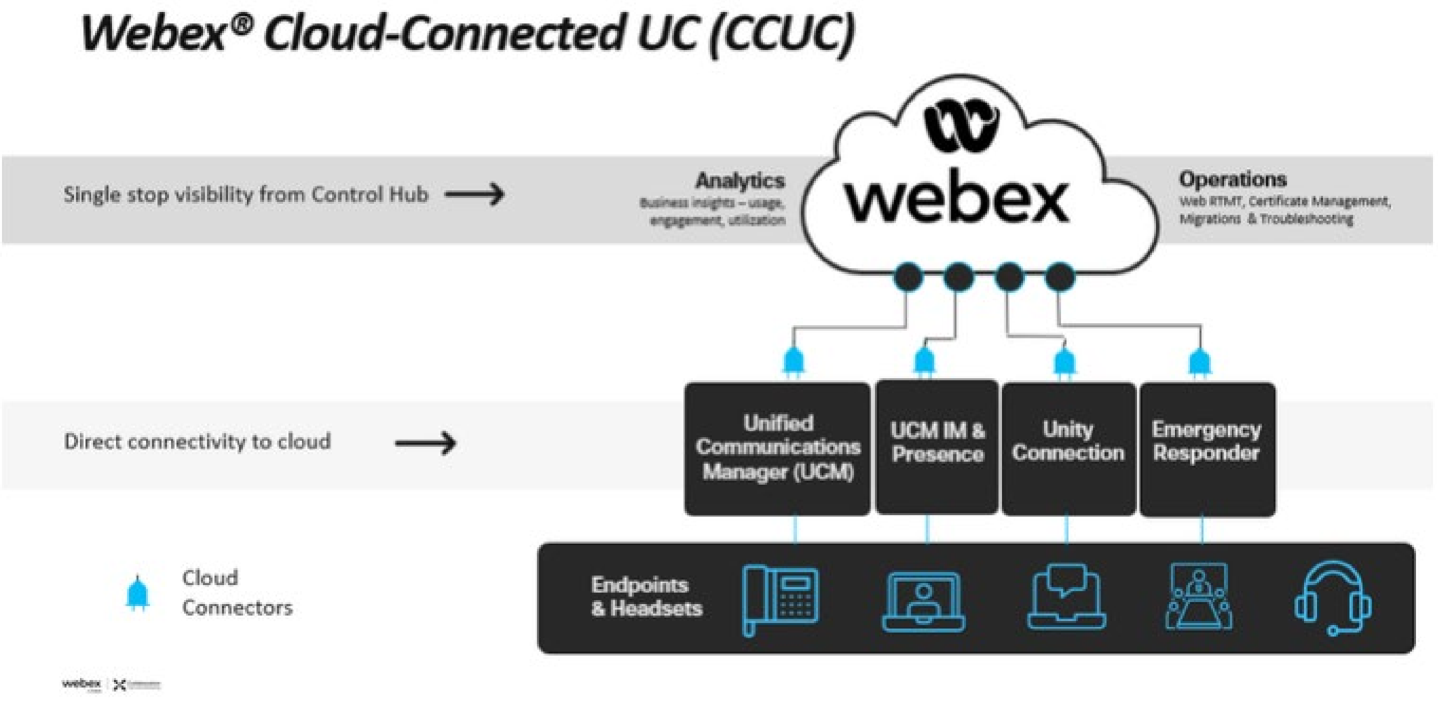
	 **Figure 1.** Full Deployment of Cisco Webex Meetings Server Behind a Firewall

System Requirements

Cisco Webex Meetings Server is compatible with Cisco UCS servers that meet or exceed the specifications presented in this section.

Module	Requirements
Host server	Cisco UCS C-Series rack server or equivalent B-Series blade server
Network interfaces	• Minimum 1 physical Network Interface Card (NIC) for a nonredundant configuration • Redundant configurations must have all NIC interfaces duplicated and connected to an independent switching fabric
Internal storage (direct attached storage [DAS]) for ESXi hosts where internal machines are deployed	Minimum of 4 drives in a RAID-10 or RAID-5 configuration
Internal storage (DAS)for ESXi hosts where Internet Reverse Proxy (IRP) virtual machines are deployed	Minimum of 2 drives in a RAID-1 configuration
SAN storage	Can be used as a substitute for DAS
Network-Attached Storage (NAS)	Can be used as a substitute for DAS or SAN
Hypervisor	• ESXi versions and vSphere licenses • 1 VMware license per processor socket
Email server	• Fully Qualified Domain Name (FQDN) of the mail server that the system uses to send emails • Port number: Default value of the Simple Mail Transfer Protocol (SMTP) port number is 25 or 465 (secure SMTP port number)

	(https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings-server/datasheet-c78-717754.html.)  (https://community.cisco.com/t5/collaboration-knowledge-base/cisco-webex-meetings-overview/ta-p/3648888.)

	 (https://community.cisco.com/t5/collaboration-knowledge-base/webex-cloud-connected-uc-connect-your-cucm-to-webex-cloud-for/ta-p/4172103.)

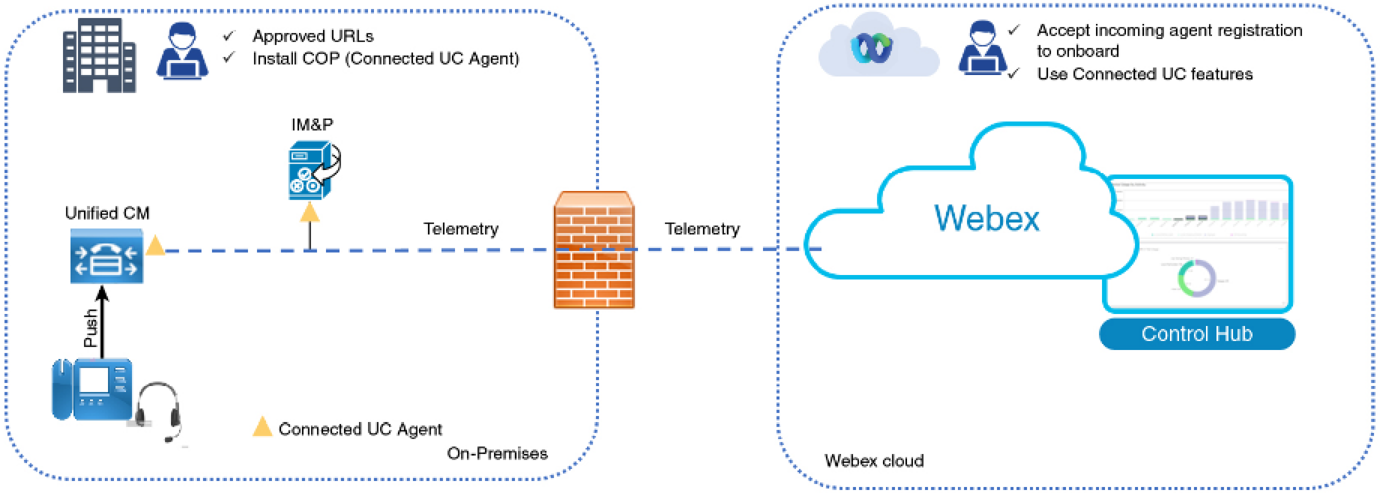
	 *Cloud-Connected UC Deployment* (https://help.webex.com/en-us/article/jv0u1db/Webex-Cloud-Connected-UC-Overview.)

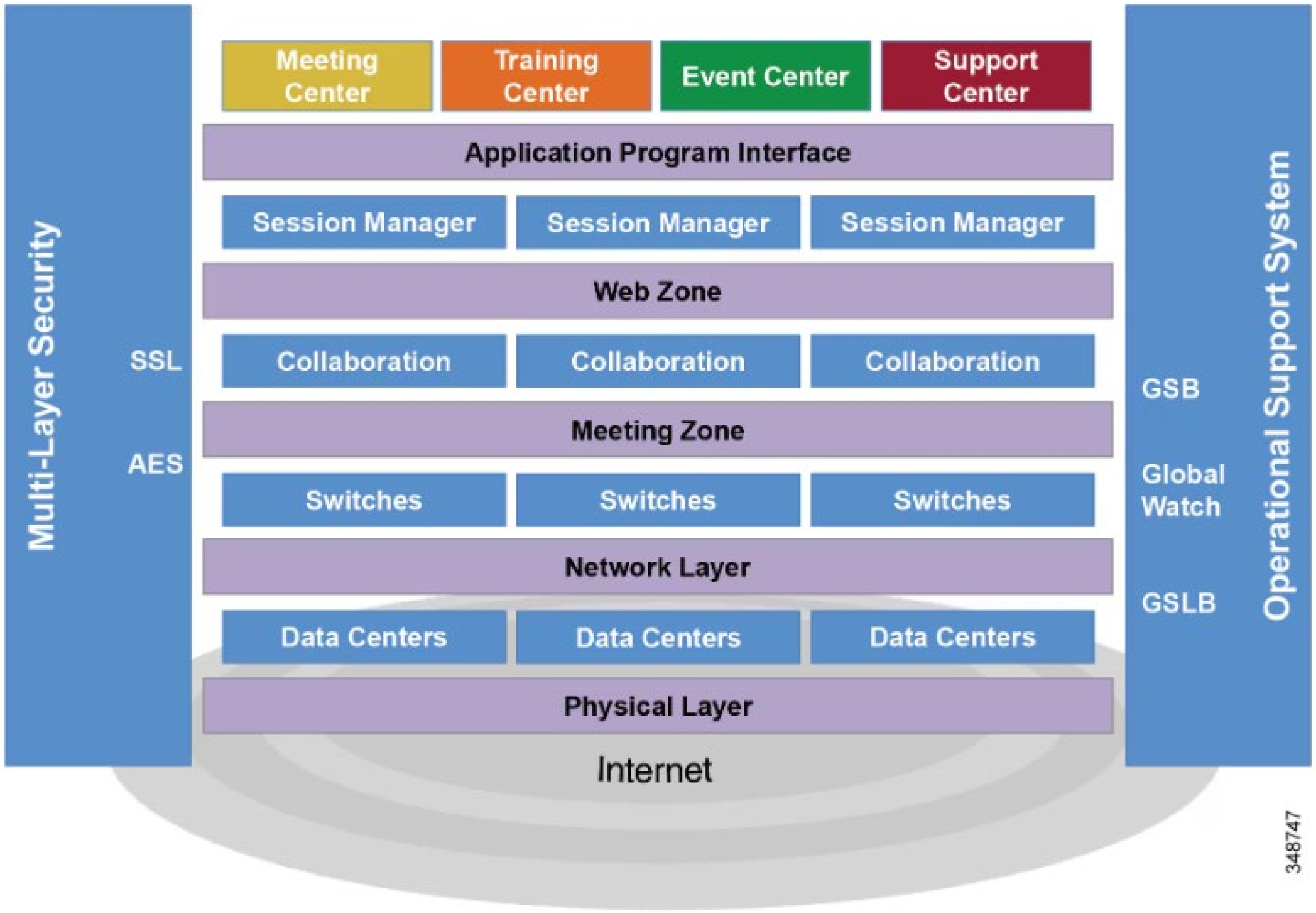
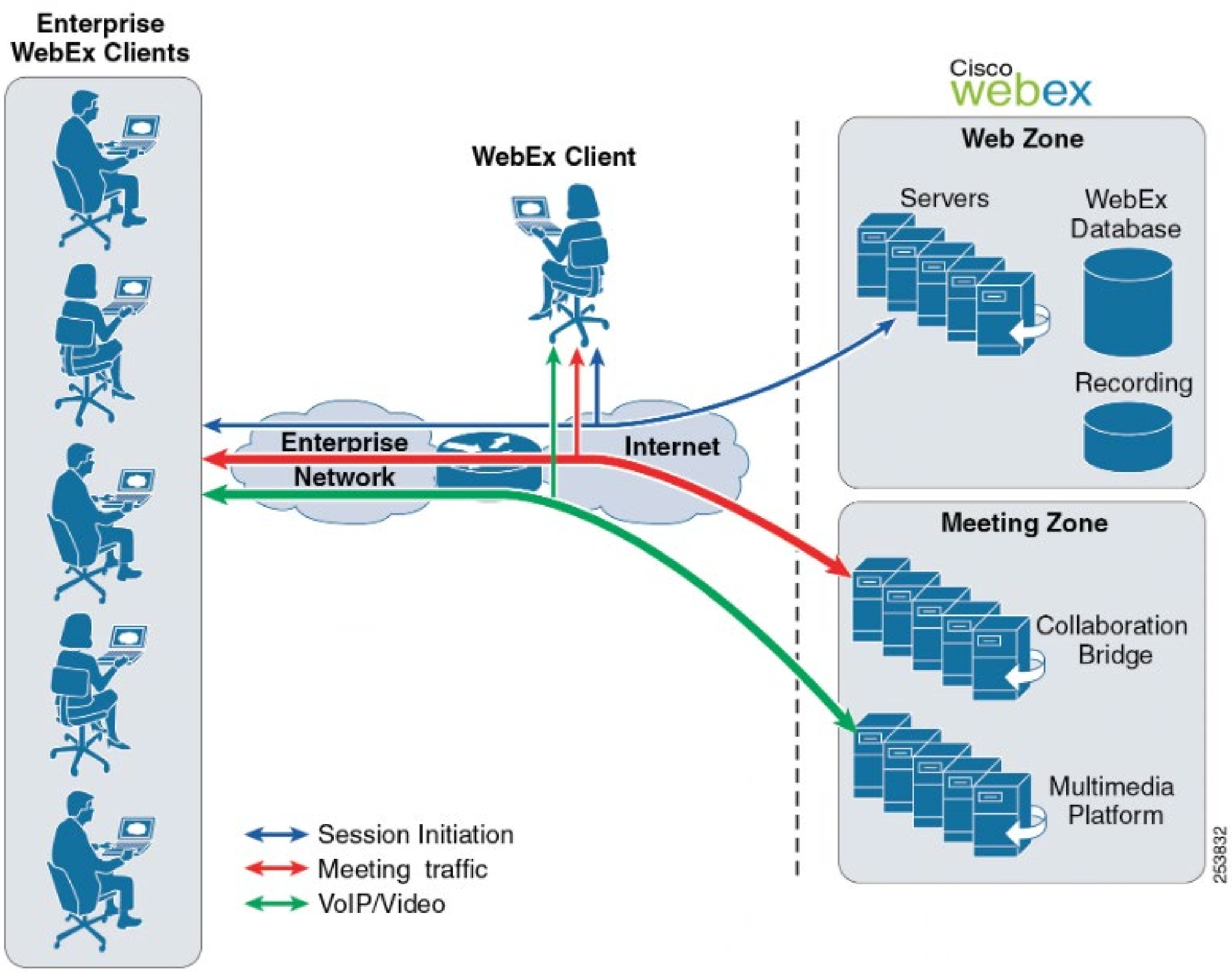
	Figure 11-14 Cisco WebEx Collaboration Cloud Architecture 

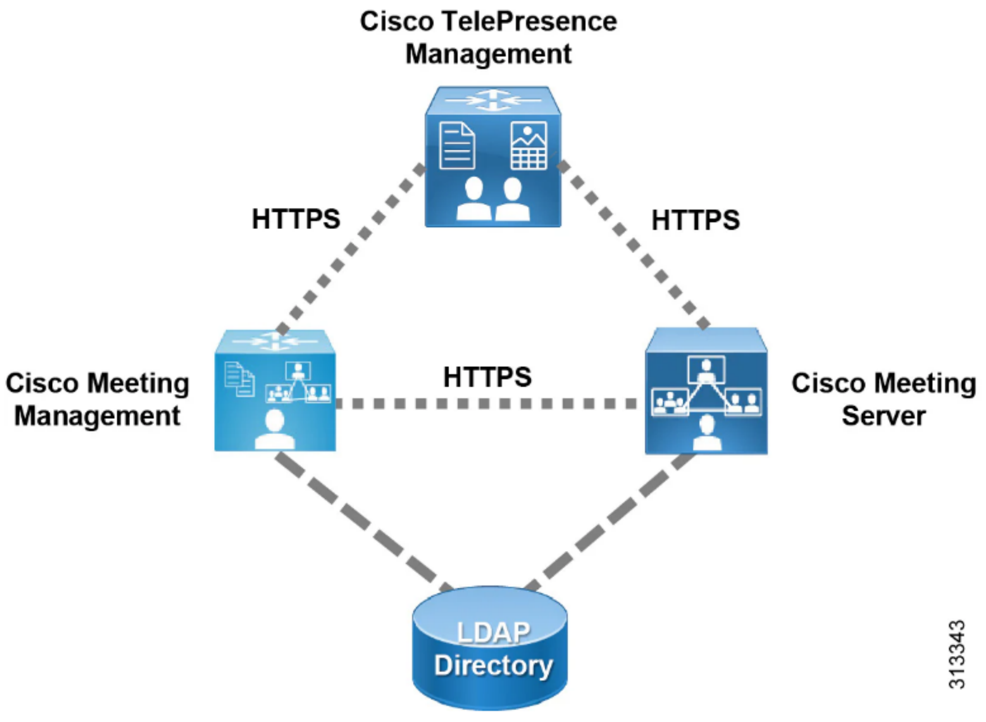
Figure 11-15 WebEx Deployment



	Cisco Rich Media Conferencing consists of the conferencing solutions described below. The details pertaining to each solution are described in each individual section that follows. • Cisco Unified CM Audio Conferencing This solution allows Unified CM to use its internal software component or external hardware digital signal processors (DSPs) as the resources to perform audio conferencing. • Cisco Meeting Server Cisco Meeting Server is an on-premises video conferencing solution. Each user has a personal Space that can be used to conduct meetings. Users can manage items such Space creation, adding members to a Space, and PIN creation from the Cisco Meeting App. • Cisco Collaboration Meeting Rooms Hybrid Cisco CMR Hybrid combines the on-premises video conference and the WebEx Meeting Center conference into a single meeting, which allows TelePresence and WebEx participants to join and share voice, video, and content. CMR Hybrid meetings can be either scheduled or non-scheduled. • Cisco WebEx Meeting Center Video Conferencing Cisco WebEx Meeting Center Video Conferencing (formerly Cisco Collaboration Meeting Rooms (CMR) Cloud) is an alternate conferencing deployment model that does not require any on-premises conferencing resources or management infrastructure. It supports both scheduled and non-scheduled meetings as well as TelePresence, audio, and WebEx participants in a single call, all hosted in the cloud. • Cisco WebEx Meetings Server Where cloud-based web and audio conferencing is not suitable, it is possible to use the on-premises WebEx Meetings Server solution. This product offers a standalone audio, video, and collaboration web conferencing platform. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab12/collab12/confernc.html.)

	Network Traffic Planning Network traffic planning for Cisco WebEx Meeting Center Video Conferencing consists of the following elements: • WebEx Clients bandwidth The WebEx meeting client uses Scalable Video Coding (SVC) technology to send and receive video. It uses multi-layer frames to send video, and the receiving client automatically selects the best possible resolution to receive video that typically requires 1.2 to 3 Mbps of available bandwidth. For more information regarding network traffic planning for WebEx clients, refer to the *Cisco WebEx Network Bandwidth* white paper, available at https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meeting-center/white_paper_c11-691351.html • Bandwidth for video device from enterprise to WebEx Cloud For optimal SIP audio and video quality, Cisco recommends setting up the video bandwidth for at least 1.5 Mbps per device screen in the region associated with the endpoint registering with Cisco Unified CM. For example, if a triple-screen device registers with Unified CM, video bandwidth of 4.5 Mbps should be allocated in the associated region. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab12/collab12/confernc.html.)

Figure 3-3 Cisco Meeting Management Architecture



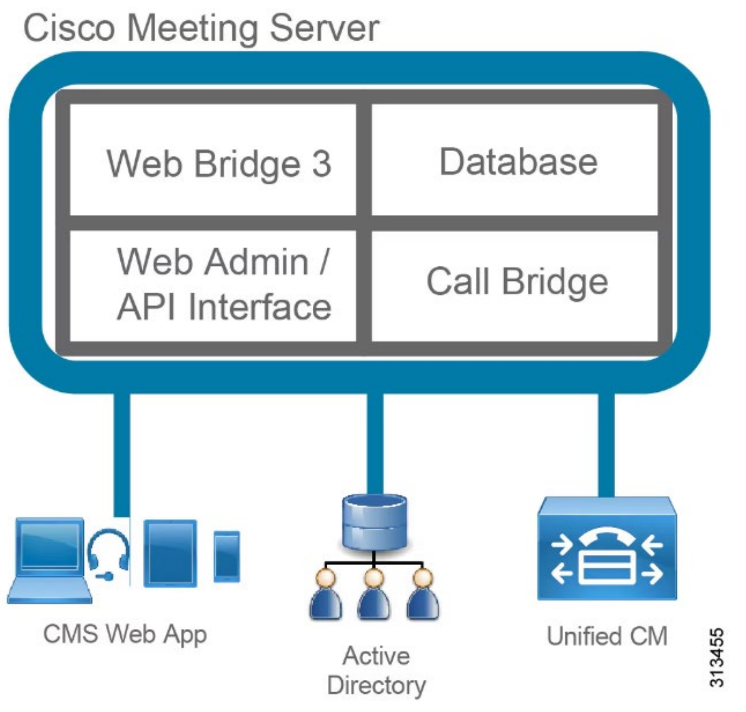
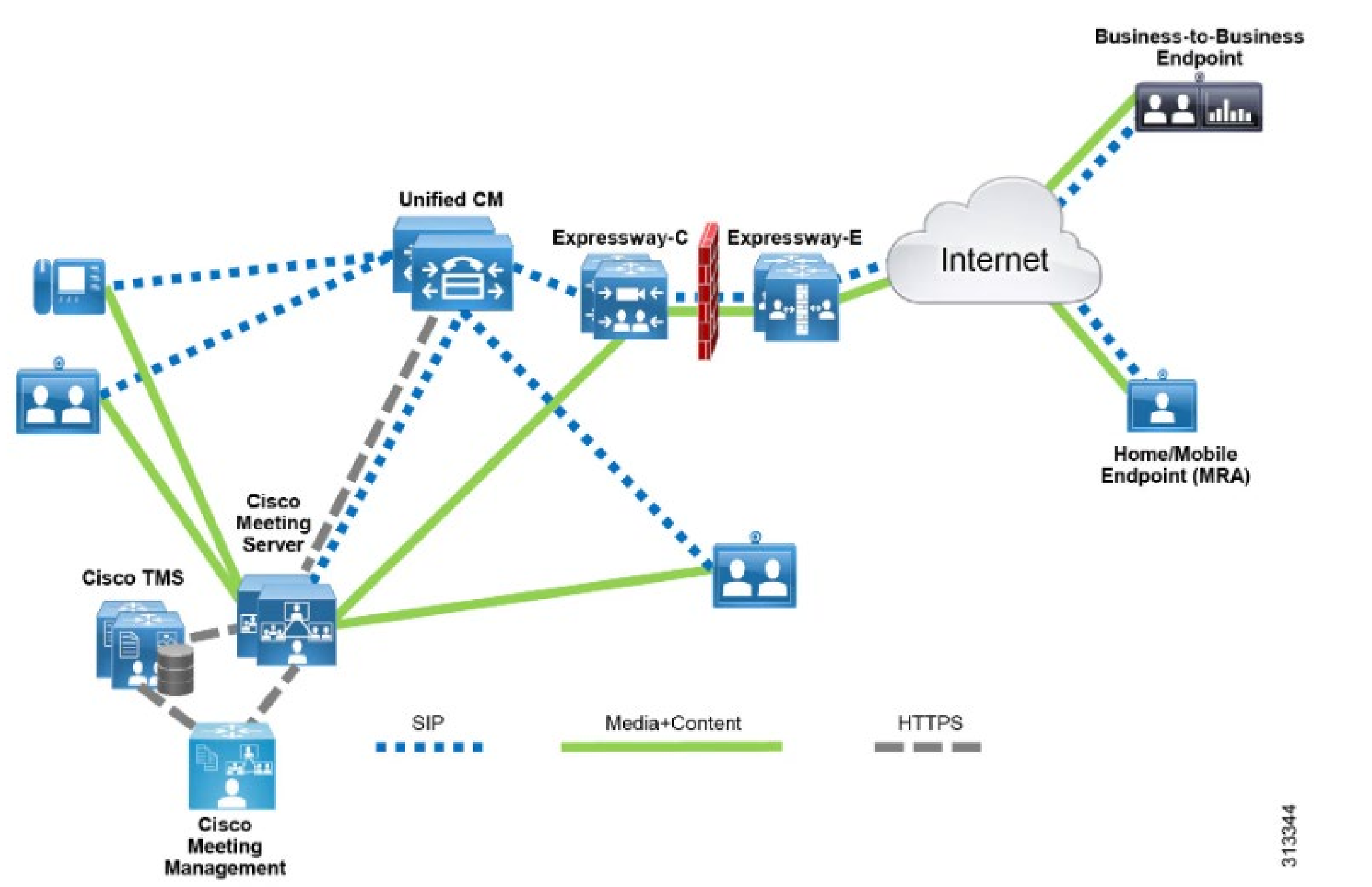
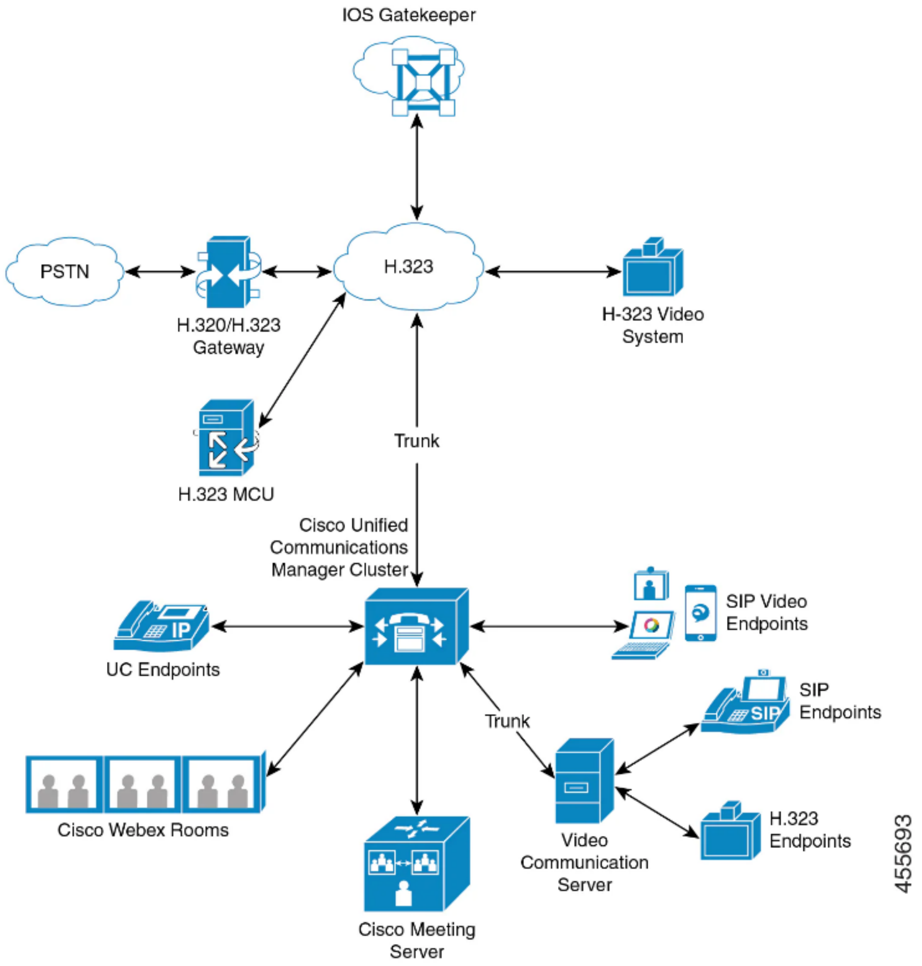
	 Cisco Meeting Server Web Bridge 3 Database Web Admin / API Interface Call Bridge CMS Web App Active Directory Unified CM 313455

	Figure 3-5 Standard Deployment  (https://www.cisco.com/c/en/us/td/docs/solutions/CVD/Collaboration/enterprise/14/collbcvd/conferencing.html.)

	How to deploy Cisco Meeting Server Talk with a Cisco salesperson or partner to learn about deployment and choose the best options for you. Select a platform Cisco Meeting Server software has been optimized to run on our UCS-based Cisco Meeting Server 1000 and Cisco Meeting Server 2000 Choose a licensing option Our multiparty option supports per-meeting licensing. Or you can purchase capacity units, as in a traditional license model. Consider add-on features Include recording ports, or consider Solution Plus partner Vbrick for recording/streaming distribution and Vyopta for assurance and analytics. Download the software Get the Cisco Meeting App for Macs and PCs on our site or from iTunes. Use the Apple Store for iOS devices. Download now › Go to iTunes › Platforms Cisco Meeting Server 1000 This Cisco UCS x86 server supports up to 120 simultaneous HD video conferencing calls. Cisco Meeting Server 2000 This Cisco UCS x86 server supports up to 875 simultaneous HD video conferencing calls. (https://www.cisco.com/c/en/us/products/conferencing/meeting-server/index.html.)
a memory; and a processor, wherein the processor executes instructions stored in the memory to cause the scalable video coding router to:	The Accused Instrumentalities include a memory and a processor, wherein the processor executes instructions stored in the memory to cause the scalable video coding router to [perform the claimed steps]. For example:

Video Network

The following illustration provides an example of a video network that uses a single Unified Communications Manager cluster. In a successful video network, any endpoint can call any other endpoint. Video availability only exists if both endpoints are video-enabled. Video capabilities extend across trunks.



	The Cisco video conference portfolio comprises the following video bridges: • Cisco TelePresence MCU series • Webex Meeting Server (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/14SU2/cucm_b_feature-configuration-guide-for-cisco14su2/cucm_m_video-telephony.html.)

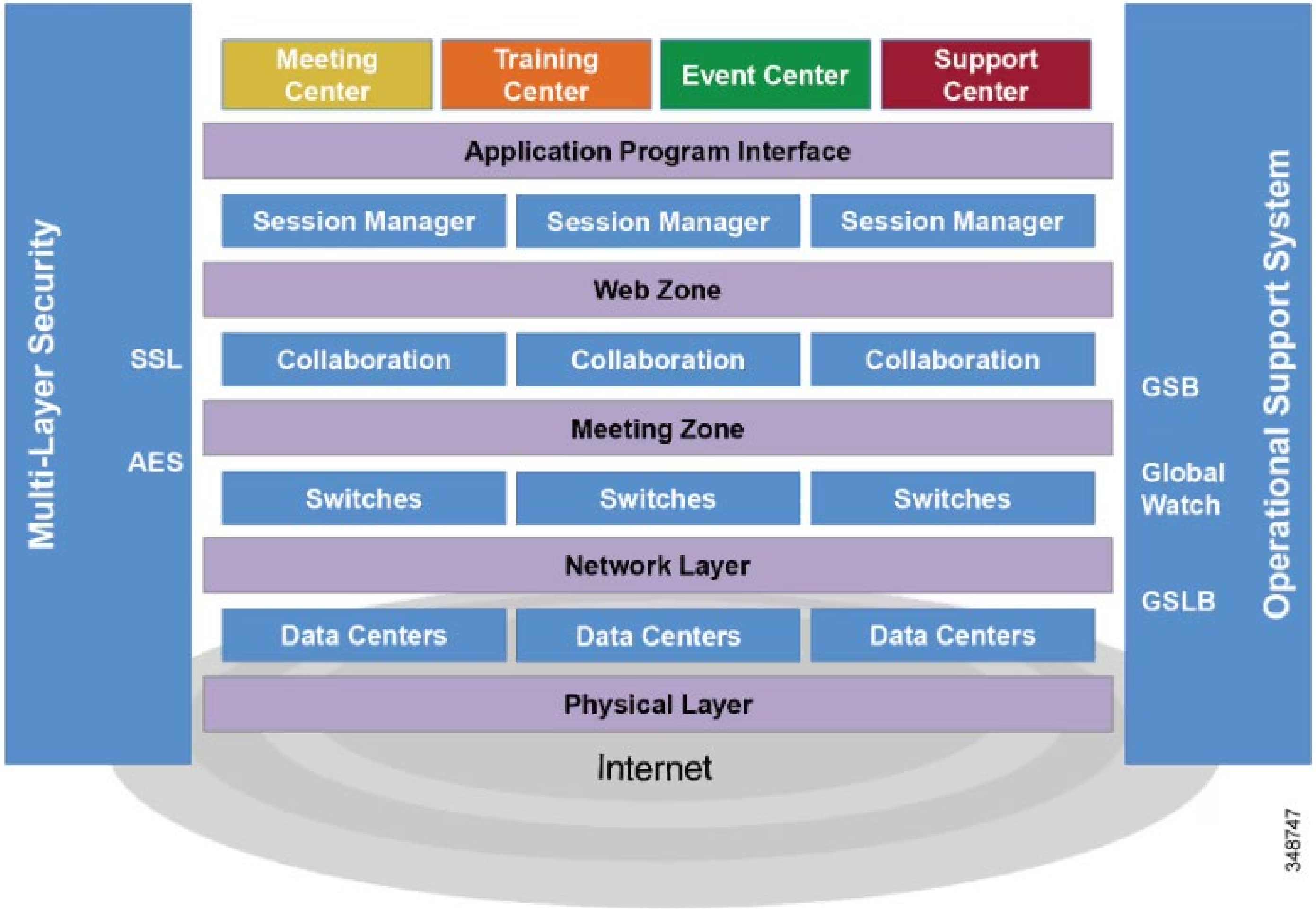
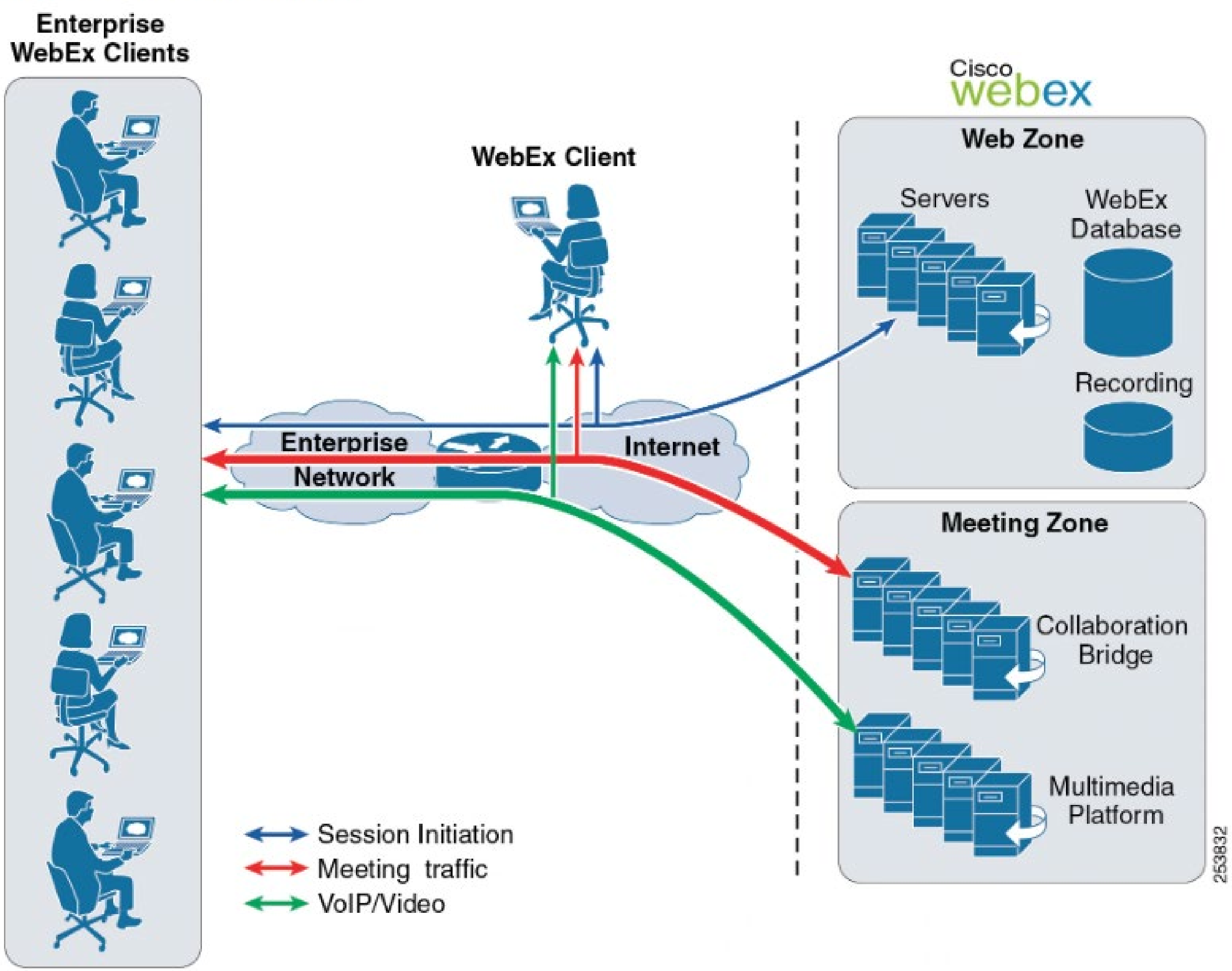
	Figure 11-14 Cisco WebEx Collaboration Cloud Architecture 

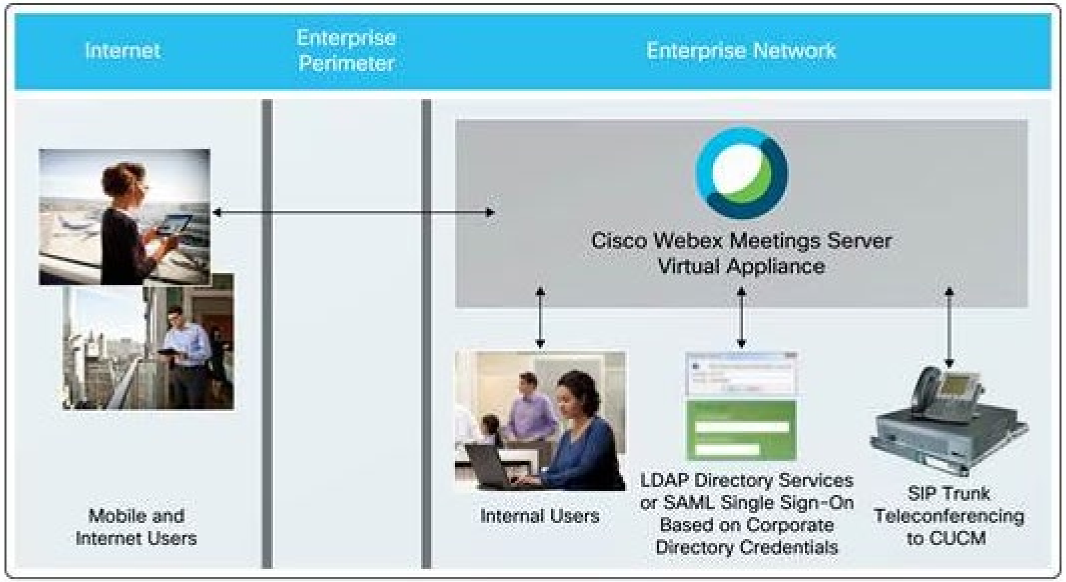
Figure 11-15 WebEx Deployment



(https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab12/collab12/confernc.html.)

Product Overview

Cisco Webex Meetings Server is a virtualized, software-based solution that runs on Cisco Unified Computing System™ (Cisco UCS®) servers and VMware. It uses virtual appliance technology for rapid turn-up of services to end users. With Cisco Webex Meetings Server, there are two options for enabling mobile users to more securely access Cisco Webex conferences without going through a VPN. The first option is to deploy reverse proxy (or edge servers) in the enterprise perimeter (or DMZ). The second option, shown in Figure 1, is to deploy the reverse proxy servers behind your internal firewall, thus eliminating all DMZ components and related information security concerns.



Optimized for 100% Secure, Behind-the-Firewall VPN-Less Access That Integrates with Your Corporate User Management and UC Infrastructure

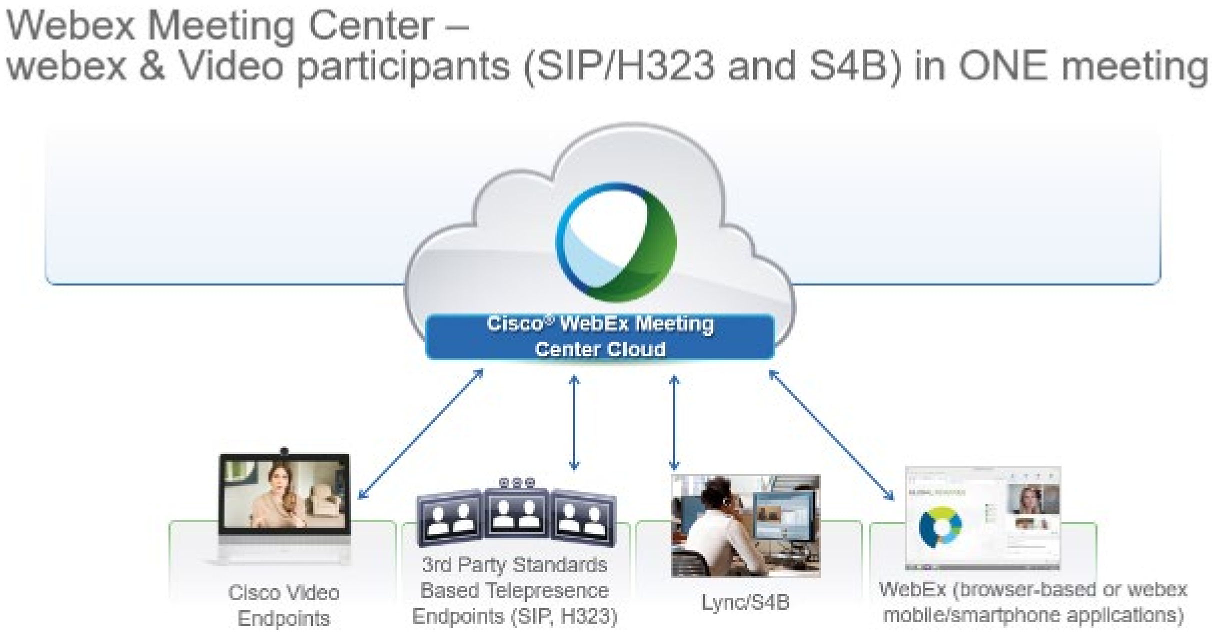
Figure 1.

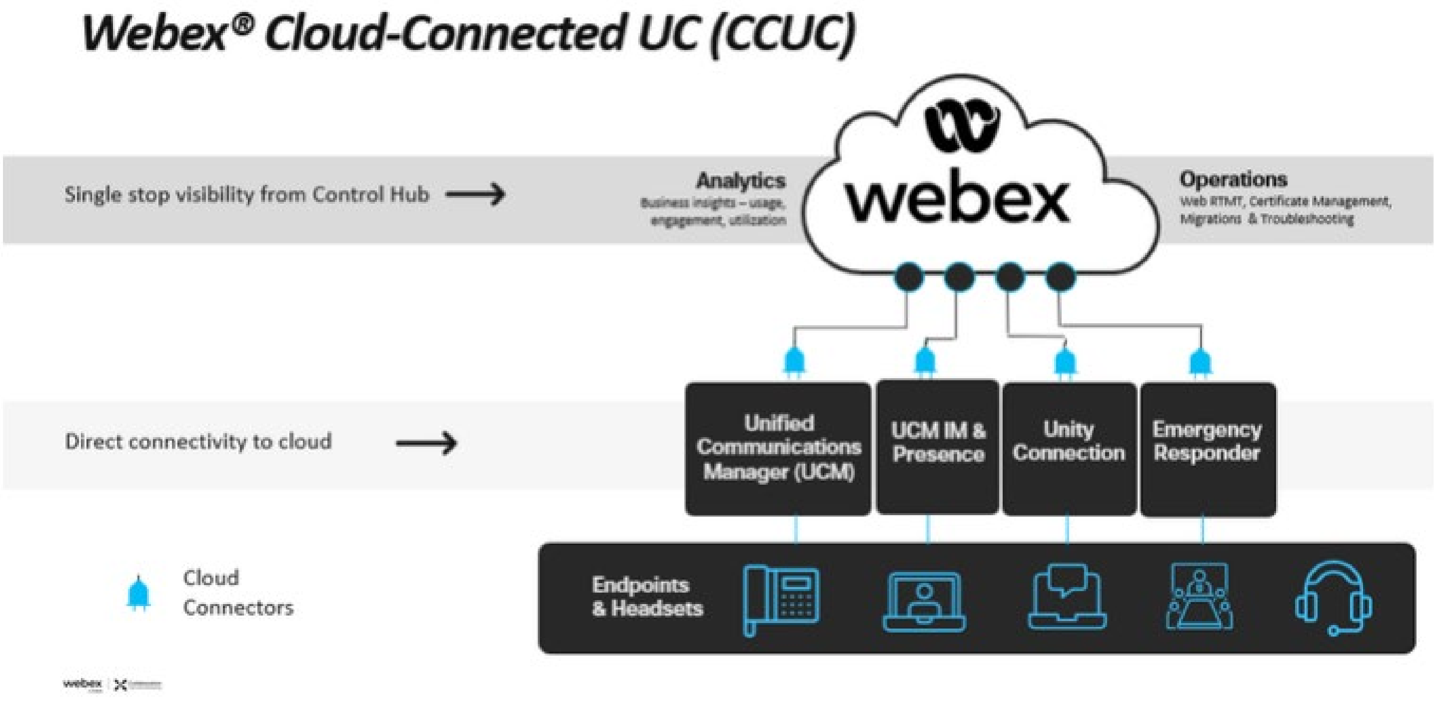
Full Deployment of Cisco Webex Meetings Server Behind a Firewall

System Requirements

Cisco Webex Meetings Server is compatible with Cisco UCS servers that meet or exceed the specifications presented in this section.

Module	Requirements
Host server	Cisco UCS C-Series rack server or equivalent B-Series blade server
Network interfaces	• Minimum 1 physical Network Interface Card (NIC) for a nonredundant configuration • Redundant configurations must have all NIC interfaces duplicated and connected to an independent switching fabric
Internal storage (direct attached storage [DAS]) for ESXi hosts where internal machines are deployed	Minimum of 4 drives in a RAID-10 or RAID-5 configuration
Internal storage (DAS)for ESXi hosts where Internet Reverse Proxy (IRP) virtual machines are deployed	Minimum of 2 drives in a RAID-1 configuration
SAN storage	Can be used as a substitute for DAS
Network-Attached Storage (NAS)	Can be used as a substitute for DAS or SAN
Hypervisor	• ESXi versions and vSphere licenses • 1 VMware license per processor socket
Email server	• Fully Qualified Domain Name (FQDN) of the mail server that the system uses to send emails • Port number: Default value of the Simple Mail Transfer Protocol (SMTP) port number is 25 or 465 (secure SMTP port number)

	(https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings-server/datasheet-c78-717754.html.)  (https://community.cisco.com/t5/collaboration-knowledge-base/cisco-webex-meetings-overview/ta-p/3648888.)

	 (https://community.cisco.com/t5/collaboration-knowledge-base/webex-cloud-connected-uc-connect-your-cucm-to-webex-cloud-for/ta-p/4172103.)

How to deploy Cisco Meeting Server

Talk with a Cisco salesperson or partner to learn about deployment and choose the best options for you.

Select a platform

Cisco Meeting Server software has been optimized to run on our UCS-based Cisco Meeting Server 1000 and Cisco Meeting Server 2000

Choose a licensing option

Our multiparty option supports per-meeting licensing. Or you can purchase capacity units, as in a traditional license model.

Consider add-on features

Include recording ports, or consider Solution Plus partner Vbrick for recording/streaming distribution and Vyopta for assurance and analytics.

Download the software

Get the Cisco Meeting App for Macs and PCs on our site or from iTunes. Use the Apple Store for iOS devices.

Download now › Go to iTunes ›

Platforms

Cisco Meeting Server 1000

This Cisco UCS x86 server supports up to 120 simultaneous HD video conferencing calls.

Cisco Meeting Server 2000

This Cisco UCS x86 server supports up to 875 simultaneous HD video conferencing calls.

(https://www.cisco.com/c/en/us/products/conferencing/meeting-server/index.html.)

Table 3. Ordering information

Platform (step 1)	Description
CTI-CMS-1K-M6-K9	Cisco Meeting Server 1000 M6
CTI-CMS-2K-M6-K9	Cisco Meeting Server 2000 M6
R-CMS-K9	Call Bridge Activation key for a third-party server or CMS 1K, only if not using Smart Licensing
R-CMS-2K-K9	Call Bridge Activation key for CMS 2K, only if not using Smart Licensing

(https://www.cisco.com/c/en/us/products/collateral/conferencing/meeting-server/datasheet-c78-742168.html.)

For example, the Accused Instrumentalities implement the AV1 standard. The AV1 standard discloses a method for transmitting video signals (e.g., video bitstream). The AV1 standard is a video coding & decoding standard. It discloses that the AV1 standard is developed for video bitstream communication application like Internet related application, streaming, etc. It induces transmission of encoded video bitstream via Internet.

	AV1 Bitstream & Decoding Process Specification *Last modified: 2019-01-08 11:48 PT* *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 (Title Page)*, available at https://aomediacodec.github.io/av1-spec/av1-spec.pdf. Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.)

(http://dgql.org/~unlord/MHV2018.pdf.)

	Video Calls The typical video call includes two or three Real-Time Protocol (RTP) streams in each direction (that is, four or six streams). The call can include the following stream types: • Video (H.261, H.263, H.263+, H.264-SVC, X-H.264UC, H.264-AVC, H.265, AV1 and VT Camera wideband video codecs) • Far-End Camera Control (FECC) - Optional • Binary Floor Control Protocols (BFCP) **AV1 Codec Support** AV1 is a next-generation video codec developed by the Alliance for Open Media. The benefits of AV1 are: • Reduced bandwidth consumption and better visual quality by utilizing better compression efficiency compared to other video encodings • Enables video for users on very low bandwidth networks • Significant screen sharing efficiency improvements over other codecs Unified Communication Manager supports negotiation of AV1 codec to establish media if endpoints support the AV1 codec. When both endpoints support Multiple Codecs in Answer, Unified CM negotiates all the matching codecs including AV1 based on the preference order received. The endpoint will then use one of the codecs from the negotiated codec list for media streaming. In a low bandwidth environment, the AV1 codec is preferred by the endpoint over other codecs in the negotiated list. When both the endpoints involved in the call do not support the Multiple Codec in Answer, and the AV1 is the preferred codec over other codecs, Unified CM selects AV1 as the negotiated codec.

	SIP Video SIP video supports the following video calls by using the SIP Signaling Interface (SSI): • SIP to SIP • SIP to H.323 • SIP to SCCP • SIP intercluster trunk • H.323 trunk • Combination of SIP and H.323 trunk SIP video calls also provide media control functions for video conferencing. Unified Communications Manager video supports SIP on both SIP trunks and lines support video signaling. SIP supports the H.261, H.263, H.263+, H.264 (AVC), H.264 (SVC), X-H.264UC (Lync), and AV1 video codecs (it does not support the wideband video codec that the VTA uses). (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/14SU2/cucm_b_feature-configuration-guide-for-cisco14su2/cucm_m_video-telephony.html.) **AV1 Codec Support** Unified Communications Manager now supports negotiation and passthrough of AVI codec. The AV1 is a modern codec that provides better compression and hence can provide the same user experience as H.264 video codec at half the bandwidth. AV1 codec will be supported by Cisco Webex Desk Pro Endpoint, Webex Codec Pro, and Room Panorama systems. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/rel_notes/14_0_1/cucm_b_release-notes-for-cucm-imp-14_0_1/cucm_m_new-and-changed-features.html.)

webex ahead

Collaboration Workspaces Customer Experience Event Management Innovation & AI

Sear

The AV1 video codec comes to Webex!

On Dec 15, 2020 - By Webex Team | 4 Min Read

Thomas Davies & Sijia Chen – Webex is rolling out the AV1 video codec into production early next year. This will bring our media quality to the next level. As a founding member of AOM, Cisco is proud to introduce this advanced video technology into the real time communications market

It's here! We have begun the process of rolling out the advanced AV1 video codec across Webex, taking video quality to the next level in the process, and replacing the aging H.264 standard.

	What do I need to use AV1 in Webex? Transmitting AV1 is supported when sharing screens or applications with "Optimize for motion and video" selected , and when the machine you are on has at least four cores. Receiving AV1 is supported for any machine with at least two cores. AV1 will automatically be used for sharing this type of screen content whenever all participants in a meeting support it, otherwise it will automatically revert to H.264.

How we are rolling out AV1

Adopting a brand-new video codec has an impact on every part of our Collaboration portfolio, so we are going step-by-step.

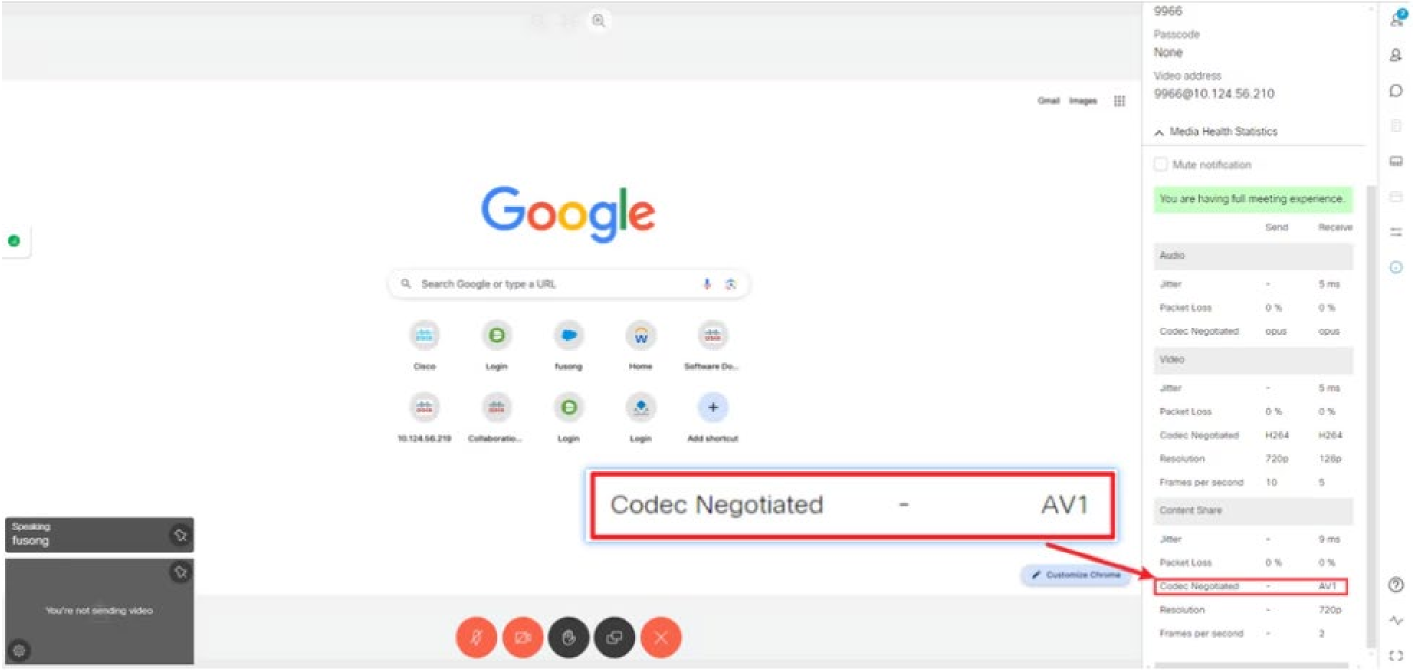
In future releases we will systematically expand where we deploy AV1. The immediate next steps are to support AV1 for other desktop share modes – either optimized for text and images, or automatically optimized. AV1 works just as well for these modes too, but we are being careful to change things gradually to make sure the user experience is perfect at each step.

Webex employs a fundamentally switched architecture, where video from each participant in a meeting is coded on their machine at different qualities and sent via a server to the other meeting participants. Initially, if some of those participants cannot support AV1 then we will automatically fall back to using H.264. Over time we will also remove these restrictions by applying ad hoc transcoding between AV1 and H.264 for those participants. This will also allow AV1 meetings to be recorded without reverting to H.264, for example.

Mobile devices will also rapidly gain hardware AV1 support, and then AV1 can be rolled out to mobile too. Although our solution is software-based and very fast on ARM as well as x86 processors, it is always better to make use of hardware codecs where possible on mobile to get the best battery life possible.

We'll also be seeking to reduce the restrictions we have placed on core count for AV1 as we continue to optimize. In fact, remarkably, our AV1 solution uses little more CPU than H.264. However, there are a huge range of different machines out there, and again we are moving gradually to safeguard user experience.

	(https://blog.webex.com/engineering/the-av1-video-codec-comes-to-webex.) "16 years after H.264, it's time for something new. Today, we demo'd an industry first: live, real-time AV1 encoding and transmission in a Webex meeting, with HD video & screen share!" — Anurag Dhingra, Cisco Webex CTO (https://medium.com/millicast/its-time-for-real-time-av1-video-encoding-withwebrtc-75a6aa64777c.) **AV1 Codec Support** Unified Communications Manager now supports negotiation and passthrough of AVI codec. The AV1 is a modern codec that provides better compression and hence can provide the same user experience as H.264 video codec at half the bandwidth. AV1 codec will be supported by Cisco Webex Desk Pro Endpoint, Webex Codec Pro, and Room Panorama systems. See the compatibility matrix for the compatible version of Webex Room devices, Cisco Expressway, and Cisco Meeting Server. (https://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/rel_notes/14_0_1/cucm_b_release-notes-for-cucm-imp-14_0_1/cucm_m_new-and-changed-features.pdf.)

	2. The media health statistics of the content receiver show the content negotiate codec is AV1 on Chrome browser when receiving the content from CMS servers.  *Receiver content codec is AV1 on chrome* (https://www.cisco.com/c/en/us/support/docs/conferencing/meeting-server/221776-configure-av1-feature-on-cms.html.)
receive a layered video data stream that comprises a base layer and a set of enhancement layers,	The Accused Instrumentalities receive a layered video data stream including a base layer and a set of enhancement layers. For example, the AV1 standard discloses receiving a layered video data stream (e.g., video bitstream) comprising a base layer (e.g., base layer) and a set of enhancement layers (e.g., enhancement layers).

	As shown below, the AV1 standard discloses an encoded video data bitstream using scalable video coding in a sequence of OBUs i.e., open bitstream unit. A metadata syntax of an OBU discloses scalability corresponding to the OBU. It discloses three types of scalabilities, Spatial scalability, Temporal scalability and Quality scalability. These scalabilities define a spatial layer having a corresponding spatial_id and a temporal layer having a corresponding temporal_id. Further, the AV1 standard discloses deriving a layered coded bitstream of base layer and enhancement layers using scalable video coding. It discloses a base layer having both spatial_id and temporal_id equal to zero and enhancement layers with either spatial_id or temporal_id equal to greater than zero values. **Bitstream** The sequence of bits generated by encoding a sequence of frames. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **5.8.1. General metadata OBU syntax**

metadata_obu( ) {	Type
metadata_type	leb128()
if (metadata_type == METADATA_TYPE_ITUT_T35)	
metadata_itut_t35( )	
else if (metadata_type == METADATA_TYPE_HDR_CLL)	
metadata_hdr_cll( )	
else if (metadata_type == METADATA_TYPE_HDR_MDCV)	
metadata_hdr_mdcv( )	
else if (metadata_type == METADATA_TYPE_SCALABILITY)	
metadata_scalability( )	
else if (metadata_type == METADATA_TYPE_TIMECODE)	
metadata_timecode( )	
}	

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 34 of 669) **5.8.5. Metadata scalability syntax** | metadata_scalability( ) { | Type | | **scalability_mode_idc** | f(8) | | if (scalability_mode_idc == SCALABILITY_SS) | | | scalability_structure( ) | | | } | | *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)*

5.8.6. Scalability structure syntax

scalability_structure( ) {	Type
spatial_layers_cnt_minus_1	f(2)
spatial_layer_dimensions_present_flag	f(1)
spatial_layer_description_present_flag	f(1)
temporal_group_description_present_flag	f(1)
scalability_structure_reserved_3bits	f(3)
if (spatial_layer_dimensions_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1 ; i++) {	
spatial_layer_max_width[i]	f(16)
spatial_layer_max_height[i]	f(16)
}	
}	
if (spatial_layer_description_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1; i++)	
spatial_layer_ref_id[i]	f(8)
}	
if (temporal_group_description_present_flag) {	
temporal_group_size	f(8)
for (i = 0; i < temporal_group_size; i++) {	
temporal_group_temporal_id[i]	f(3)
temporal_group_temporal_switching_up_point_flag[i]	f(1)
temporal_group_spatial_switching_up_point_flag[i]	f(1)
temporal_group_ref_cnt[i]	f(3)
for (j = 0; j < temporal_group_ref_cnt[i]; j++) {	
temporal_group_ref_pic_diff[i][j]	f(8)
}	

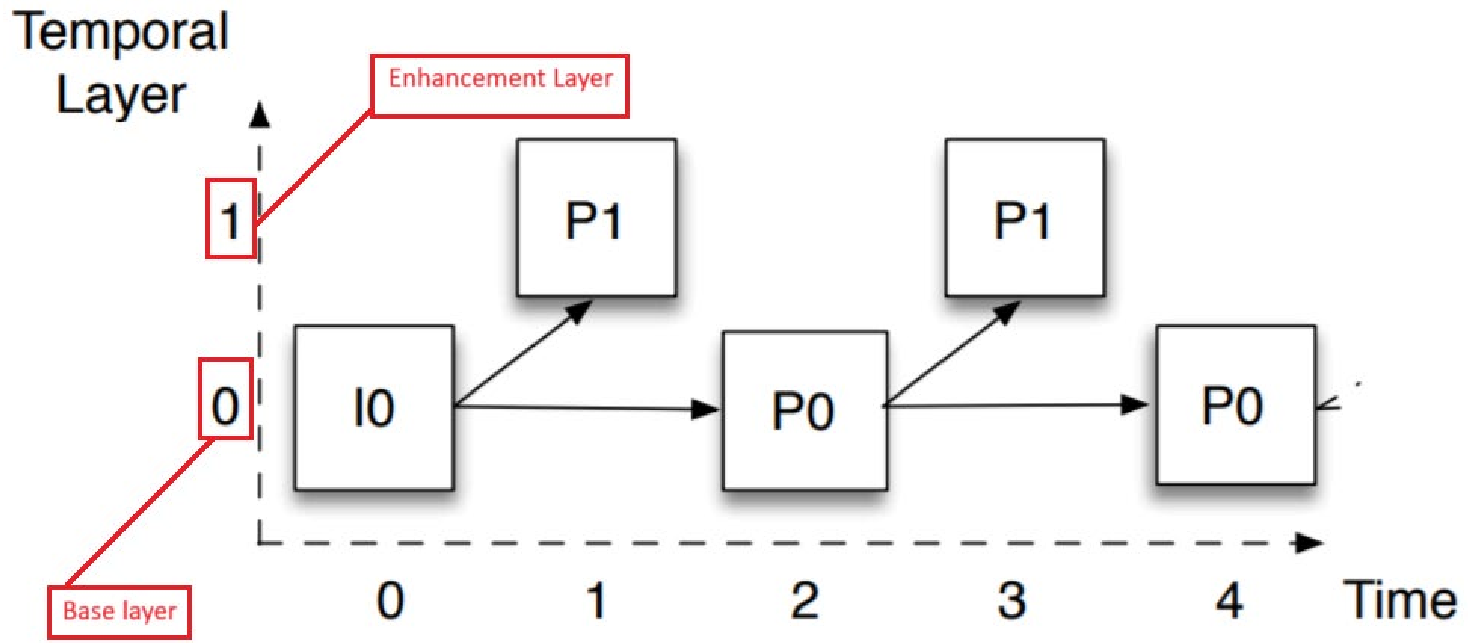
Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

spatial_layer_ref_id[i] specifies the spatial_id value of the frame within the current temporal unit that the frame of layer i uses for reference. If no frame within the current temporal unit is used for reference the value must be equal to 255.

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669) Note that for a given picture, all frames follow the same inter-picture temporal dependency structure. However, the frame rate of each layer can be different from each other. The specified dependency structure in the scalability structure data must be for the highest frame rate layer. **temporal_group_temporal_id[i]** specifies the temporal_id value for the i-th picture in the temporal group. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669)* **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)*

	The AV1 codec can maintain up to eight reference frames, of which up to seven can be referenced by any new frame. AV1 also allows a frame to use another frame of a different spatial resolution as a reference frame. This allows internal resolution changes without requiring the use of key frames. These features together enable an AV1 encoder to implement various forms of coarse-grained scalability, including temporal, spatial, and quality scalability modes, as well as combinations of these, without the need for explicit scalable coding tools. Spatial and quality layers define different and possibly dependent representations of a single input frame. For a given spatial layer, temporal layers define different frame rates of video. Spatial layers allow a frame to be encoded at different spatial resolutions, whereas quality layers allow a frame to be encoded at the same spatial resolution but at different qualities (and thus with different amounts of coding error). AV1 supports quality layers as spatial layers without any resolution changes; hereinafter, the term "spatial layer" is used to represent both spatial and quality layers. This payload format specification provides for specific mechanisms through which such temporal and spatial scalability layers can be described and communicated. Temporal and spatial scalability layers are associated with non-negative integer IDs. The lowest layer of either type has an ID equal to 0. (https://aomediacodec.github.io/av1-rtp-spec/.) **Layer** A set of tile group OBUs with identical spatial_id and identical temporal_id values. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 3 of 669)* **Base layer** The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)*

	Enhancement layer A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)* As shown below, the AV1 standard discloses a base layer and enhancement layers for a coded video bitstream. **Note:** The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* **Note:** Examples are given for non-scalable cases, but the constraints also apply to each operating point of a scalable stream. For example, consider a 60fps spatial scalable stream with a base layer at 960x540, and an enhancement layer at 1920x1080. The operating point containing just the base layer may be labelled as level 3.0, while the operating point containing both the base and enhancement layer may be labelled as level 4.1. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 641 of 669)*



Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 126 of 669)

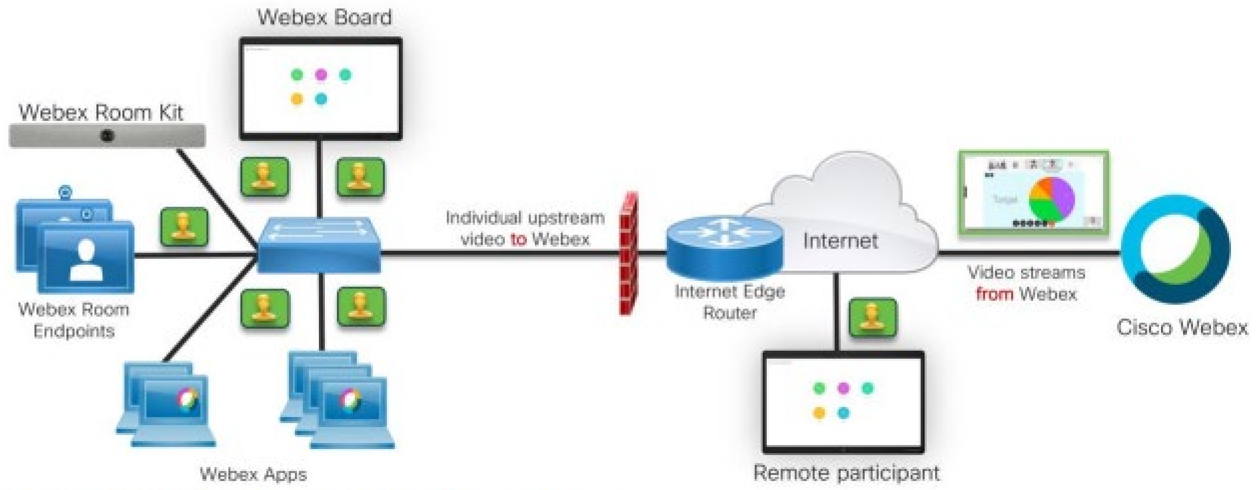
<table>
<tr>
<td></td>
<td>


Figure 11: Video Streams in a Webex Meeting

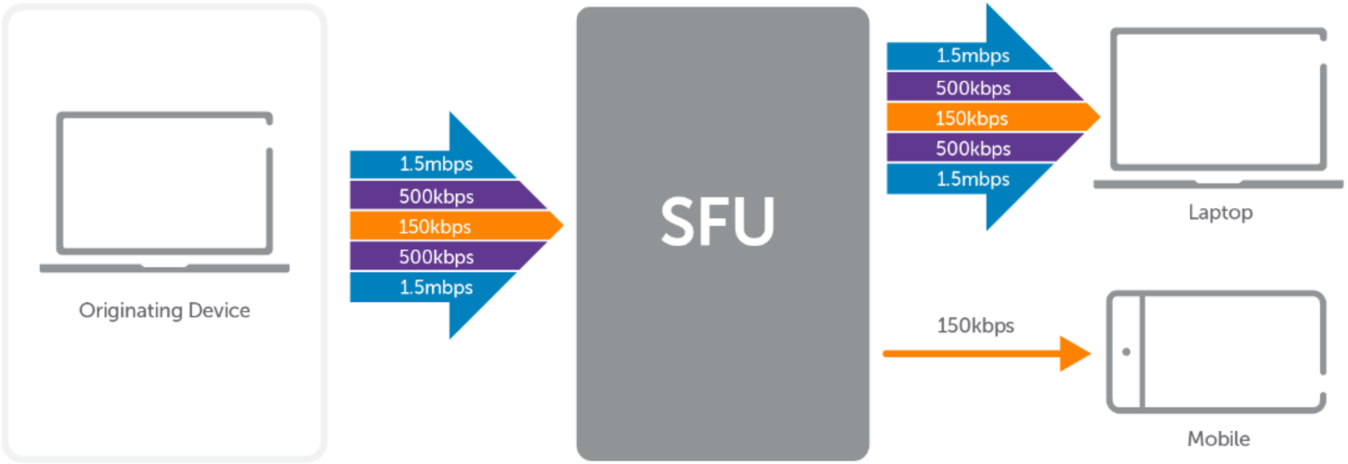
(https://www.cisco.com/c/dam/td-xml/en_us/collaboration/cloud_cmr/pcia_2_0/reports/Troubleshooting_Audio_and_Video_Quality_Using_Webex_Control_Hub.pdf.)
</td>
</tr>
<tr>
<td>identify bandwidth-limited conditions of an internet protocol network between the video router and a set of video receivers;</td>
<td>The Accused Instrumentalities identify bandwidth-limited conditions of an internet protocol network between the video router and a plurality of video receivers.

For example, the AV1 standard discloses identifying bandwidth-limited conditions (e.g., network conditions, available bandwidth condition for a receiving device, etc.) of an internet protocol network (e.g., Internet, etc.) between a video router (e.g., a video data transmitter such as a video bitstream encoder, etc.) and a plurality of video receivers (e.g., video data receivers such as a video bitstream decoder, etc.).</td>
</tr>
</table>

	Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) For example, one of the applications of the AV1 standard discloses a selective forwarding unit or middlebox (SFU or SFM) which analyses bandwidth condition between a publisher/transmitter (e.g., video router) and a playback

	device (e.g., video receiver). According to the bandwidth condition, it selectively forwards layers to the playback device. **How Does SVC Work?** Video Router 1. **Publisher**: Transmits a single media stream (with multiple video layers) to the Selective Forwarding Unit (SFU). 2. **SFU**: Examines the stream and selectively encodes it into high and low bitrate layers. It then distributes these layers based on each playback device's network limitations and available bandwidth. 3. **Playback Devices**: Receive the appropriate layers, ensuring optimal quality for each participant. Video Receiver (https://www.linkedin.com/pulse/scalable-video-coding-svc-sudeep-kumar-nofdc.)

	The SFU then sends the adjusted stream along to the appropriate **playback device**. Depending on the available bandwidth and other limitations of the target devices, some could receive a lower quality stream while others get the whole onion (so to speak).  (https://www.wowza.com/blog/scalable-video-coding-for-webrtc.) **Selective Forwarding Middlebox (SFM)** A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant.

	(https://aomediacodec.github.io/av1-rtp-spec/.) As another example, the Accused Instrumentalities are capable of identifying bandwidth-limited conditions, such as latency, jitter, and/or packet loss between backend servers and a plurality of user clients:

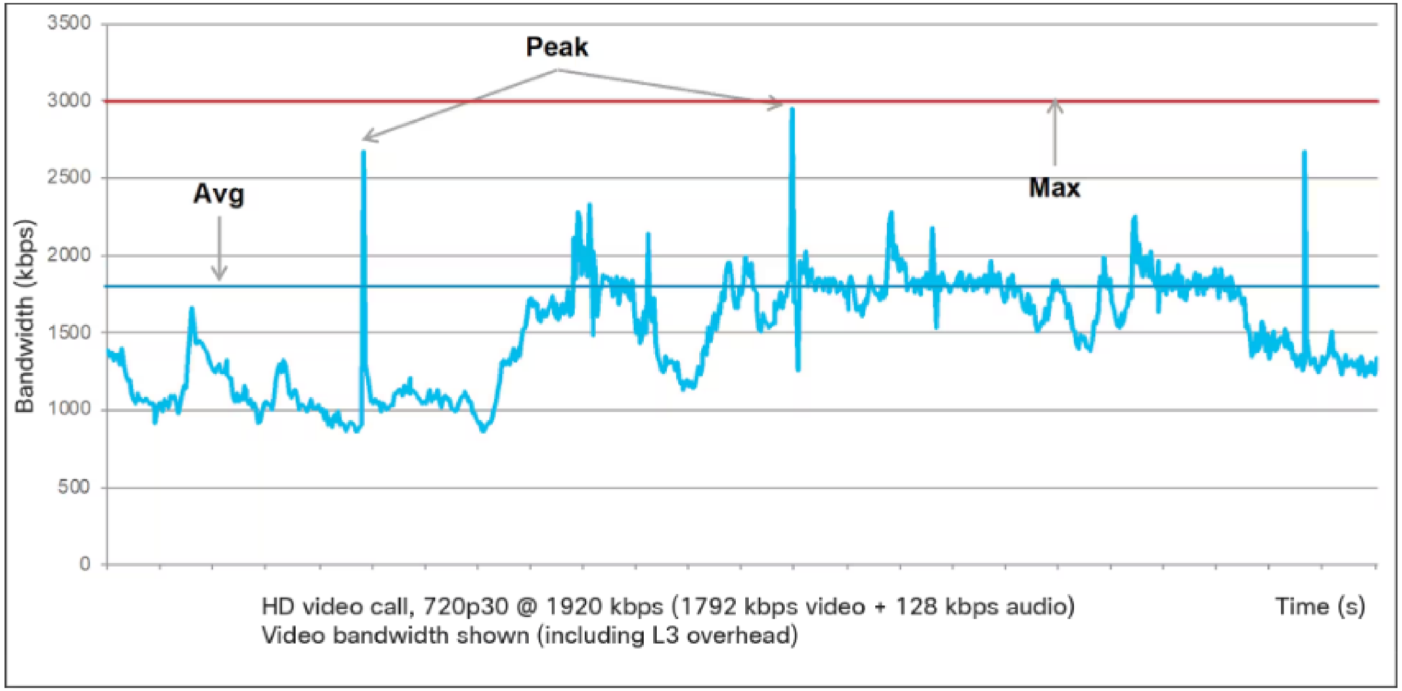
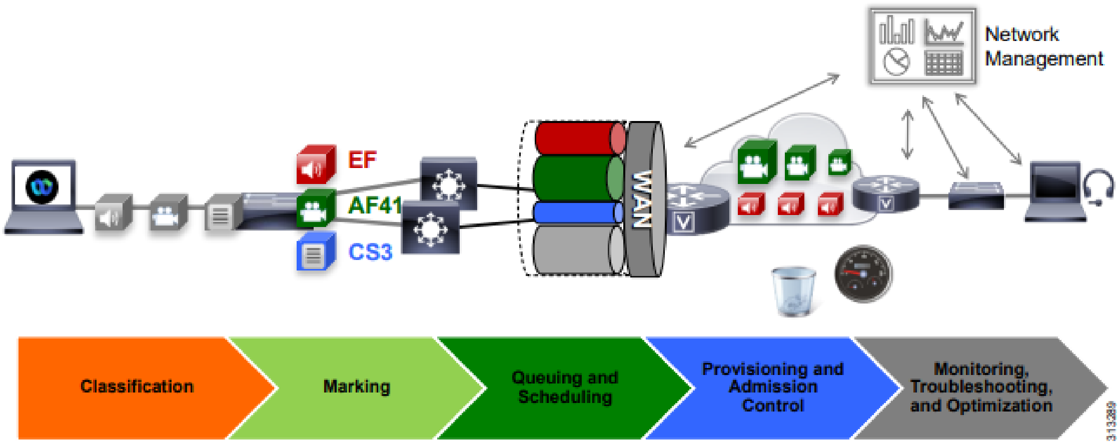
	Overview In this document we will discuss bandwidth utilization. Bandwidth values used will be in payload bit rate which does not include packetization overhead and are covered in 3 categories, average, peak and maximum bit rate: Average (**avg**) is the average over time for a meeting participant. Peak (**peak**) is the typical peak bursts over the same time period for a meeting participant. Maximum (**max**) is the maximum bit rate that the device is capable of either due to device limitations or device configuration.  **Figure 1.** Video Traffic: Bandwidth Usage High-definition Video Call (https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings/white_paper_c11-691351.html.)

Figure 1 Architecture for Bandwidth Management



Recommended Deployment

Modify the existing on-premises QoS switch and WAN and Internet policies to include Webex Services identification, classification, and marking.

- Identify and classify media and signaling traffic from the Webex App and associated workloads as well as Webex Devices.
- Media and signaling marking recommendations:
    1. Mark all audio with Expedited Forwarding class EF (includes all audio of voice-only calls as well as audio for all types of video calls).
    2. Mark all Webex App video with an Assured Forwarding class of AF41 for a prioritized video class of service and AF42 for an opportunistic class of service. The marking of AF41 or AF42 will depend on the choice of whether to deploy opportunistic video during the on-premises deployment phase.
    3. Mark all call signaling with CS3. (All call signaling in HTTPS traffic will be marked based on the enterprise's current policy of traffic marking for HTTP/HTTPS – unless using NBAR which is covered in detail below)
- Configure QoS on all media originating and terminating applications such as the Video Mesh Nodes, Expressway and Cisco Unified Border Element.
- Update the WAN edge ingress re-marking policy.
- Update the WAN edge egress queuing and scheduling policy.

Every Cisco video endpoint employs several smart media techniques to avoid network congestion, recover from packet loss, and optimize network resources. The following smart media techniques are just some of the techniques employed by Cisco Webex App and Devices:

Media resilience techniques

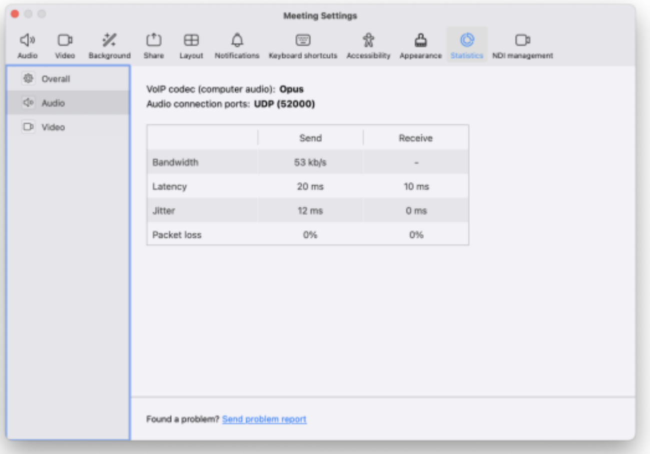
- Encoder pacing
- Forward Error Correction (FEC)
- Rate adaptation

(https://www.cisco.com/c/dam/en/us/td/docs/solutions/CVD/Collaboration/AltDesigns/BWM-Wbx.pdf.)

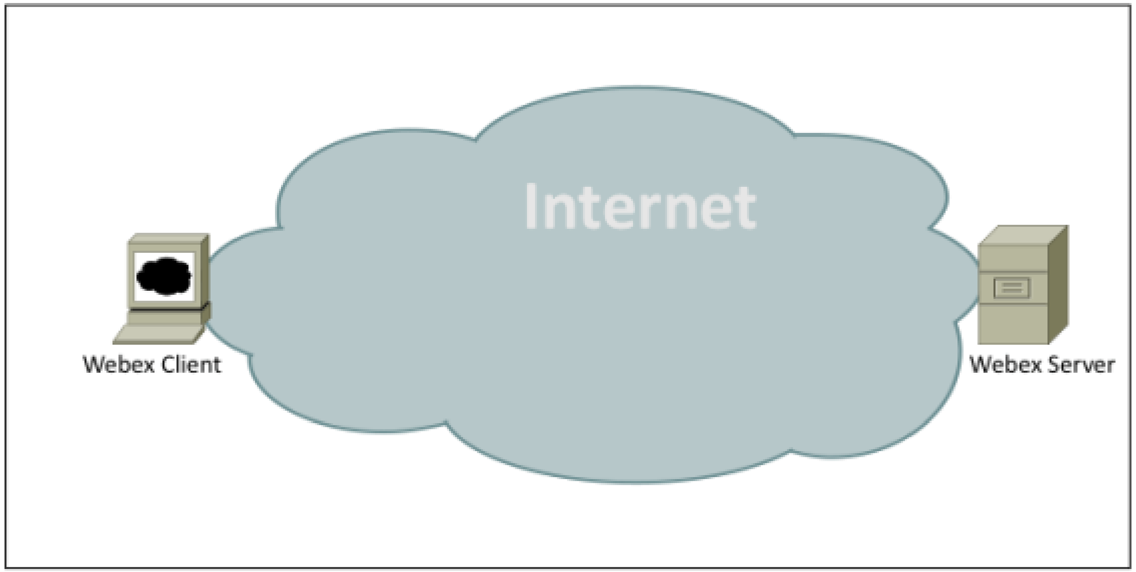
Poor Audio / Video Quality – Full-featured Meetings

Help > Health Checker > Audio and Video Statistics...

- Indicates TCP or UDP w/ Source Port
- Latency / Packet Loss / Jitter



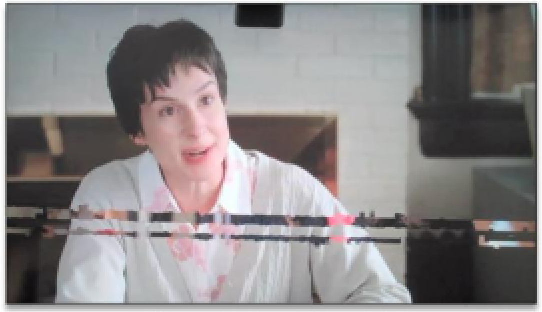
(https://www.ciscolive.com/c/dam/r/ciscolive/global-event/docs/2024/pdf/BRKCOL-3431.pdf.)

	There are three main factors that impact the quality of an audio or video call. These factors are **packet loss**, **latency**, and **jitter**. As shown in Figure 2, **packet loss is simply losing one or more packets within a stream of packets**. In this example, a Voice over IP (VoIP) packet is lost going from the Webex client to the Webex server. Loss can occur for a number of reasons in a network but most commonly it is caused by congestion or resource contention in the network or on the endpoints. This leads to routers, switches, or the endpoints themselves dropping or delaying packets. In other cases, packets encounter such a high delay that they are received too late to be played out. These packets will also be counted as lost.  ***Figure 2:*** *Packet Loss Example*

***Note**: Control Hub Diagnostics highlights potential problems with packet loss and delay by coloring voice and video streams as Good, Fair, or Poor. It is important to understand that just because Control Hub Diagnostics flags part of a conversation as Fair or Poor, it does not necessarily mean that the user had a bad experience. Voice and video compensation mechanisms and algorithms in combination with other factors can mitigate the impact of loss and delay to the point where it is not noticeable by the user.*

Video Quality Artifacts

The first step in troubleshooting video quality is to clearly understand the symptom experienced by the end user. Getting a screenshot or a picture of the display screen usually gives more information than a verbal description of the problem. The type of video artifacts observed in a screenshot or picture can be used to determine the troubleshooting path. For example, in Figure 13, pictures (a) and (b) are artifacts that are caused when video streams experience packet loss. Just by seeing the picture, you can immediately start troubleshooting to identify the source of the packet loss.



a. Video with line stripe artifacts

b. Frozen video with block artifacts

Webex Control Hub makes it quite easy to find the device type and operating system version used by the end user to join a Webex meeting. All you need is the end user's email address and meeting time to find the right meeting and to get end user's device details. Figure 14 shows the list of meetings attended by a participant with the email address, rtpmsuser1@gmail.com.

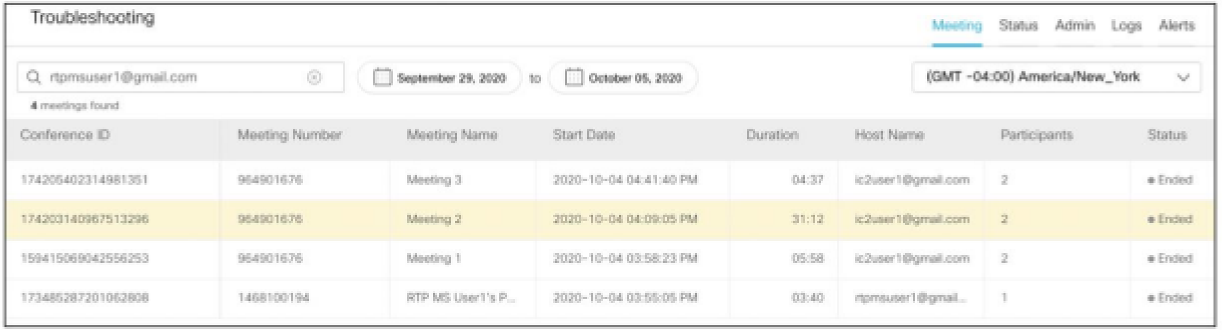


***Figure 14**: Participant Device Details*

(https://www.cisco.com/c/dam/td-xml/en_us/collaboration/cloud_cmr/pcia_2_0/reports/Troubleshooting_Audio_and_Video_Quality_Using_Webex_Control_Hub.pdf.)

	To support more than four video streams across a distribution link, it is recommended that the bandwidth of the link be set to greater than 2Mbps. Use the API or the Web Admin Interface to set the bandwidth. If using the API, PUT a value for the `peerLinkBitRate` parameter to the API object `/system/configuration/cluster`; the value will be the maximum media bit rate to use on distribution links between Call Bridges in the cluster. Alternatively, using the Web Admin Interface, go to **Configuration > Cluster > Call Bridge identity** and enter the **Peer link bit rate**. (https://www.cisco.com/c/dam/en/us/td/docs/conferencing/ciscoMeetingServer/Deployment_Guide/Version-3-9/Cisco-Meeting-Server-3-9-Scalable-and-Resilient-Deployment.pdf.)
forward the base layer from the video router to at least two of the set of video receivers;	The Accused Instrumentalities forward the base layer from the video router to at least two of the set of video receivers. For example, the AV1 standard discloses forwarding (e.g., selective forwarding using SFU – selective forwarding unit or SFM – selective forwarding middlebox) the base layer (e.g., base layer) to at least two of the plurality of video receivers (e.g., video data receiver such as a video bitstream decoder, etc.) via the internet protocol network (e.g., Internet, etc.). The AV1 standard is a video coding & decoding standard. It discloses that the AV1 standard is developed for video bitstream communication application like Internet related application, streaming, etc. It induces transmission of encoded video bitstream via Internet. It also discloses scaling according to varying bandwidth condition. Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) For example, one of the applications of the AV1 standard discloses a selective forwarding unit or middlebox (SFU or SFM) which analyses bandwidth condition between a publisher/transmitter (e.g., video router) and a playback

device (e.g., video receiver). According to the bandwidth condition, it selectively forwards layers to the playback device.

As shown below, it forwards only base layer to devices which are having low bandwidth available and other devices depending on their bandwidth conditions could receive different number of enhancement layers.

How Does SVC Work?

Video Router

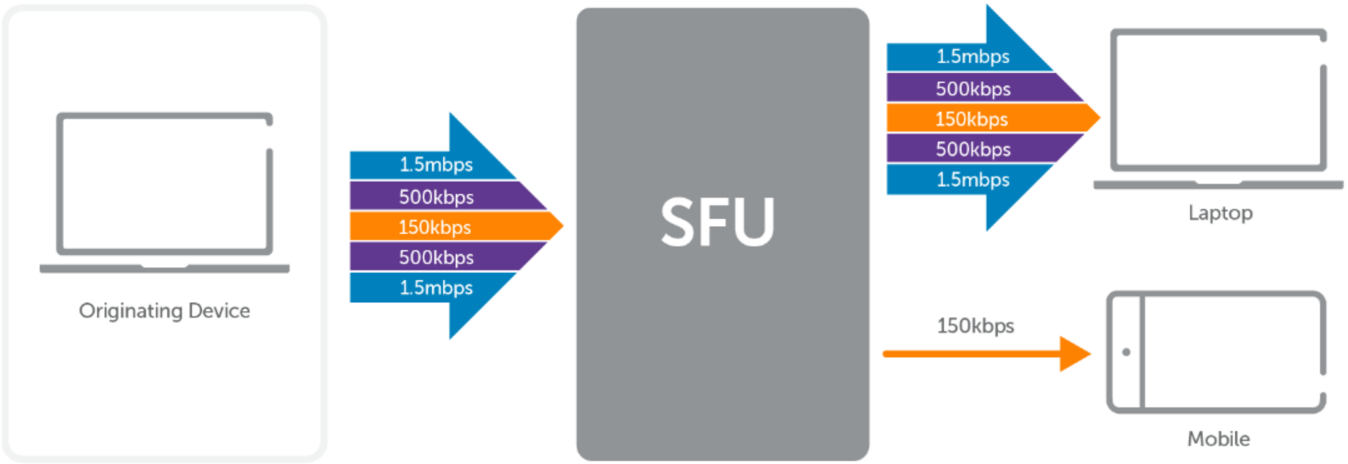
1. **Publisher**: Transmits a single media stream (with multiple video layers) to the Selective Forwarding Unit (SFU).

2. **SFU**: Examines the stream and selectively encodes it into high and low bitrate layers. It then distributes these layers based on each playback device's network limitations and available bandwidth.

3. **Playback Devices**: Receive the appropriate layers, ensuring optimal quality for each participant.

Video Receiver

(https://www.linkedin.com/pulse/scalable-video-coding-svc-sudeep-kumar-nofdc.)

	The SFU then sends the adjusted stream along to the appropriate **playback device**. Depending on the available bandwidth and other limitations of the target devices, some could receive a lower quality stream while others get the whole onion (so to speak).  (https://www.wowza.com/blog/scalable-video-coding-for-webrtc.)

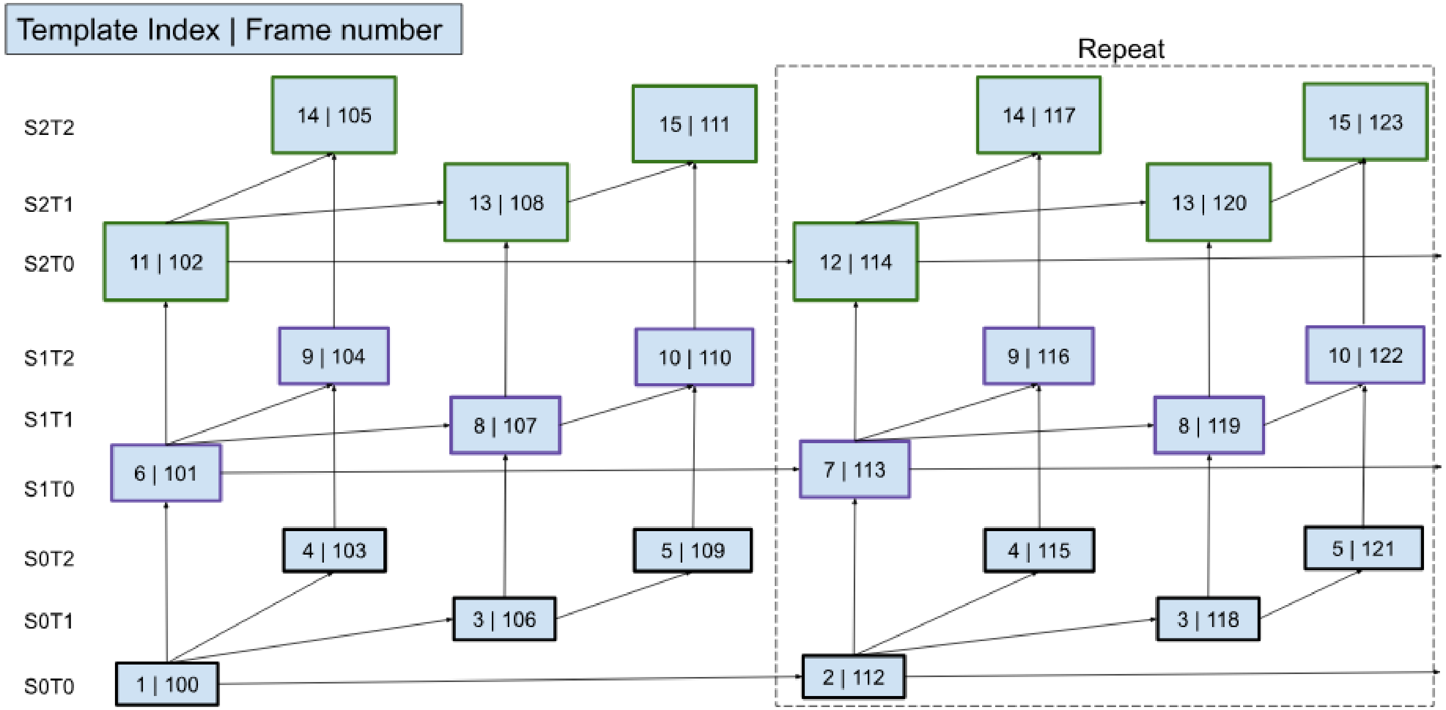
	Selective Forwarding Middlebox (SFM) A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant. ## Dependency Descriptor RTP Header Extension ### A.1 Introduction This appendix describes the Dependency Descriptor (DD) RTP Header extension. The DD is used for conveying dependency information about individual video frames in a scalable video stream. The DD includes provisions for both temporal and spatial scalability. In the DD, the smallest unit for which dependencies are described is an RTP frame. An RTP frame contains one complete coded video frame and may also contain additional information (e.g., metadata). Each RTP frame is identified by a frame_number. When spatial scalability is used, there may be multiple RTP frames produced for the same time instant. Further, this specification allows for the transmission of an RTP frame over multiple packets. RTP frame aggregation is explicitly disallowed. Hereinafter, RTP frame will be referred to as frame. The DD uses the concept of Decode targets, each of which represents a subset of a scalable video stream necessary to decode the stream at a certain temporal and spatial fidelity. A frame may be associated with several Decode targets. This concept is used to facilitate selective forwarding, as is done by a Selective Forwarding Middlebox (SFM). Typically an SFM would select one Decode target for each endpoint, and forward all frames associated with that target.

	# 6 MANE and SFM Behavior If a packet contains an OBU with an OBU extension header then the entire packet is considered associated with the layer identified by the temporal_id and spatial_id combination that are indicated in the extension header. If a packet does not contain any OBU with an OBU extension header, then it is considered to be associated with all operating points. The general function of a MANE or SFM is to selectively forward packets to receivers. To make forwarding decisions a MANE may inspect the media payload, so that it may need to be able to parse the AV1 bitstream and if so, cannot function when end-to-end encryption is enabled. An SFM does not parse the AV1 bitstream and therefore needs to obtain the information relevant to selective forwarding by other means, such as the Dependency Descriptor described in Appendix A. In addition to enabling bitstream-independent forwarding and support for end-to-end encryption, the Dependency Descriptor also enables forwarding where the metadata OBU provided in the AV1 bitstream is not sufficient to express the structure of the stream. ## 6.1.1 Example In this example, it is desired to send three simulcast encodings, each containing three temporal layers. When sending all encodings on a single SSRC, scalability mode 'S3T3' would be indicated within the scalability metadata OBU, and the Dependency Descriptor describes the dependency structure of all encodings. When sending each simulcast encoding on a distinct SSRC, the scalability mode 'L1T3' would be indicated within the scalability metadata OBU of each bitstream, and the Dependency Descriptor in each stream describes only the dependency structure for that individual encoding. A distinct spatial_id (e.g. 0, 1, 2) could be used for each stream (if a single AV1 encoder is used to produce the three simulcast encodings), but if distinct AV1 encoders are used, the spatial_id values may not be distinct.

	Indication	Description	SFM behavior
DT0	Not present	F5 is not associated with DT0.	Do not forward F5 to a DT0 client.
DT1	Discardable	No frame in DT1 will reference F5.	Should forward F5 to a DT1 client, but may discard it, e.g., when bandwidth is low.
DT2	Switch	If it is possible to decode F5, then all future frames in DT2 are also decodable.	Forward F5 to the DT2 client. In addition, the SFM can rely on the fact that if the F5 frame is decodable for a particular client then that client may be switched to DT2.
DT3	Required	Future frames in DT3 reference F5.	Forward F5 to a DT3 client. No additional decisions can be made.

A.10.2.2 L3T3 Full SVC

This example uses three Chains. Chain 0 includes frames with spatial ID equal to 0 and temporal ID equal to 0. Chain 1 includes frames with spatial ID equal to 0 or 1 and temporal ID equal to 0. Chain 2 includes all frames with temporal ID equal to 0.



Idx	S	T	Fdiffs	Chains			Decode Targets								
				0	1	2	HD30 fps	HD15 fps	HD7.5 fps	VGA30 fps	VGA15 fps	VGA7.5fps	QVGA30 fps	QVGA15 fps	QVGA7.5 fps
1	0	0		0	0	0	S	S	S	S	S	S	S	S	S
2	0	0	12	12	11	10	R	R	R	R	R	R	S	S	S
3	0	1	6	6	5	4	R	R	-	R	R	-	S	D	-
4	0	2	3	3	2	1	R	-	-	R	-	-	D	-	-
5	0	2	3	9	8	7	R	-	-	R	-	-	D	-	-
6	1	0	1	1	1	1	S	S	S	S	S	S	-	-	-
7	1	0	12,1	1	1	1	R	R	R	S	S	S	-	-	-
8	1	1	6,1	7	6	5	R	R	-	S	D	-	-	-	-
9	1	2	3,1	4	3	2	R	-	-	D	-	-	-	-	-
10	1	2	3,1	10	9	8	R	-	-	D	-	-	-	-	-
11	2	0	1	2	1	1	S	S	S	-	-	-	-	-	-
12	2	0	12,1	2	1	1	S	S	S	-	-	-	-	-	-
13	2	1	6,1	8	7	6	S	D	-	-	-	-	-	-	-
14	2	2	3,1	5	4	3	D	-	-	-	-	-	-	-	-
15	2	2	3,1	11	10	9	D	-	-	-	-	-	-	-	-
decode_target_protected_by							2	2	2	1	1	1	0	0	0

(https://aomediacodec.github.io/av1-rtp-spec/.)

	Base layer The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **Enhancement layer** A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)* 6.2.3. OBU extension header semantics **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)*

	Note: The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). Coded video data of a temporal level with temporal_id T and spatial level with spatial_id S are only allowed to reference previously coded video data of temporal_id T' and spatial_id S', where T' <= T and S' <= S. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* If a coded video sequence contains at least one enhancement layer (OBUs with spatial_id greater than 0 or temporal_id greater than 0) then all frame headers and tile group OBUs associated with base (spatial_id equals 0 and temporal_id equals 0) and enhancement layer (spatial_id greater than 0 or temporal_id greater than 0) data must include the OBU extension header. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 205 of 669)*
wherein the scalable video coding router forwards all of the set of enhancement layers to at least two of the video receivers in the set of video receivers with bandwidth-sufficient conditions,	The Accused Instrumentalities forward all of the set of enhancement layers to at least two of the video receivers in the set of video receivers with bandwidth-sufficient conditions. For example, the AV1 standard discloses forwarding (e.g., selective forwarding using SFU – selective forwarding unit or SFM – selective forwarding middlebox) all of the set of enhancement layers (e.g., enhancement layers) to at least two of the video receivers (e.g., video data receiver such as a video bitstream decoder, etc.) in the set of video receivers with bandwidth-sufficient conditions.

	The AV1 standard is a video coding & decoding standard. It discloses that the AV1 standard is developed for video bitstream communication application like Internet related application, streaming, etc. It induces transmission of encoded video bitstream via Internet. It also discloses scaling according to varying bandwidth condition. Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

AV1 Features

ROYALTY-FREE

Interoperable and open

UBIQUITOUS

Scales to any modern device at any bandwith

FLEXIBLE

For use in both commercial and non-commercial content, including user-generated content

30% BETTER COMPRESSION *

Uses less data while delivering 4k UHD video and beyond when compared to alternatives

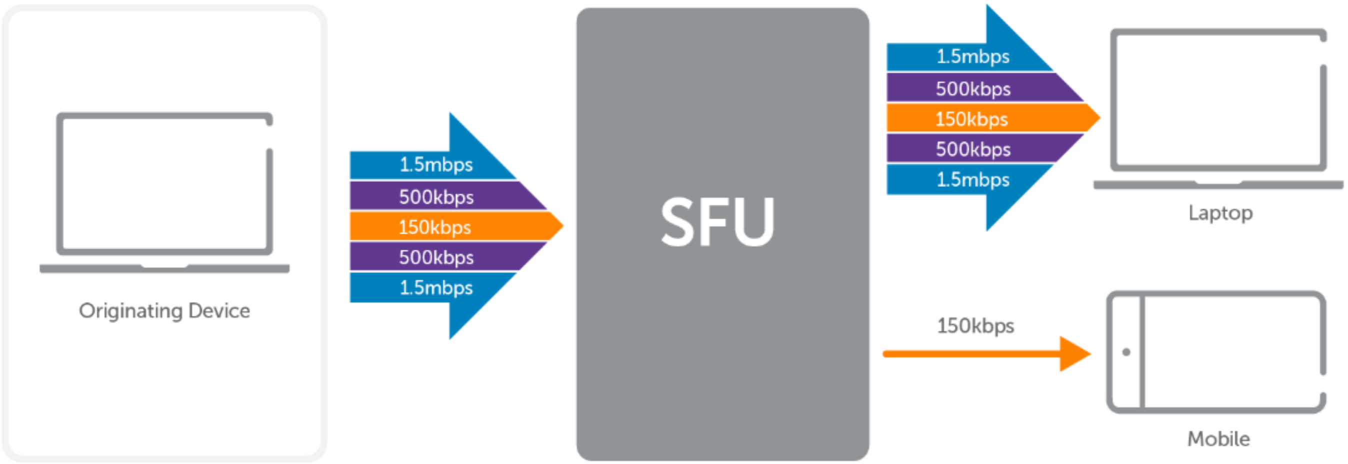
OPTIMIZED

Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services

(https://aomedia.org/av1-features/.)

For example, one of the applications of the AV1 standard discloses a selective forwarding unit or middlebox (SFU or SFM) which analyses bandwidth condition between a publisher/transmitter (e.g., video router) and a playback

	device (e.g., video receiver). According to the bandwidth condition, it selectively forwards layers to the playback device. As shown below, it forwards only base layer to devices which are having low bandwidth available and other devices depending on their bandwidth conditions could receive different number of enhancement layers. **How Does SVC Work?** Video Router 1. **Publisher**: Transmits a single media stream (with multiple video layers) to the Selective Forwarding Unit (SFU). 2. **SFU**: Examines the stream and selectively encodes it into high and low bitrate layers. It then distributes these layers based on each playback device's network limitations and available bandwidth. 3. **Playback Devices**: Receive the appropriate layers, ensuring optimal quality for each participant. Video Receiver (https://www.linkedin.com/pulse/scalable-video-coding-svc-sudeep-kumar-nofdc.)

	The SFU then sends the adjusted stream along to the appropriate **playback device**. Depending on the available bandwidth and other limitations of the target devices, some could receive a lower quality stream while others get the whole onion (so to speak).  (https://www.wowza.com/blog/scalable-video-coding-for-webrtc.)

Selective Forwarding Middlebox (SFM)
A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667).

Temporal unit
Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant.

Dependency Descriptor RTP Header Extension

A.1 Introduction

This appendix describes the Dependency Descriptor (DD) RTP Header extension. The DD is used for conveying dependency information about individual video frames in a scalable video stream. The DD includes provisions for both temporal and spatial scalability.

In the DD, the smallest unit for which dependencies are described is an RTP frame. An RTP frame contains one complete coded video frame and may also contain additional information (e.g., metadata). Each RTP frame is identified by a frame_number. When spatial scalability is used, there may be multiple RTP frames produced for the same time instant. Further, this specification allows for the transmission of an RTP frame over multiple packets. RTP frame aggregation is explicitly disallowed. Hereinafter, RTP frame will be referred to as frame.

The DD uses the concept of Decode targets, each of which represents a subset of a scalable video stream necessary to decode the stream at a certain temporal and spatial fidelity. A frame may be associated with several Decode targets. This concept is used to facilitate selective forwarding, as is done by a Selective Forwarding Middlebox (SFM). Typically an SFM would select one Decode target for each endpoint, and forward all frames associated with that target.

6 MANE and SFM Behavior

If a packet contains an OBU with an OBU extension header then the entire packet is considered associated with the layer identified by the temporal_id and spatial_id combination that are indicated in the extension header. If a packet does not contain any OBU with an OBU extension header, then it is considered to be associated with all operating points.

The general function of a MANE or SFM is to selectively forward packets to receivers. To make forwarding decisions a MANE may inspect the media payload, so that it may need to be able to parse the AV1 bitstream and if so, cannot function when end-to-end encryption is enabled. An SFM does not parse the AV1 bitstream and therefore needs to obtain the information relevant to selective forwarding by other means, such as the Dependency Descriptor described in Appendix A. In addition to enabling bitstream-independent forwarding and support for end-to-end encryption, the Dependency Descriptor also enables forwarding where the metadata OBU provided in the AV1 bitstream is not sufficient to express the structure of the stream.

6.1.1 Example

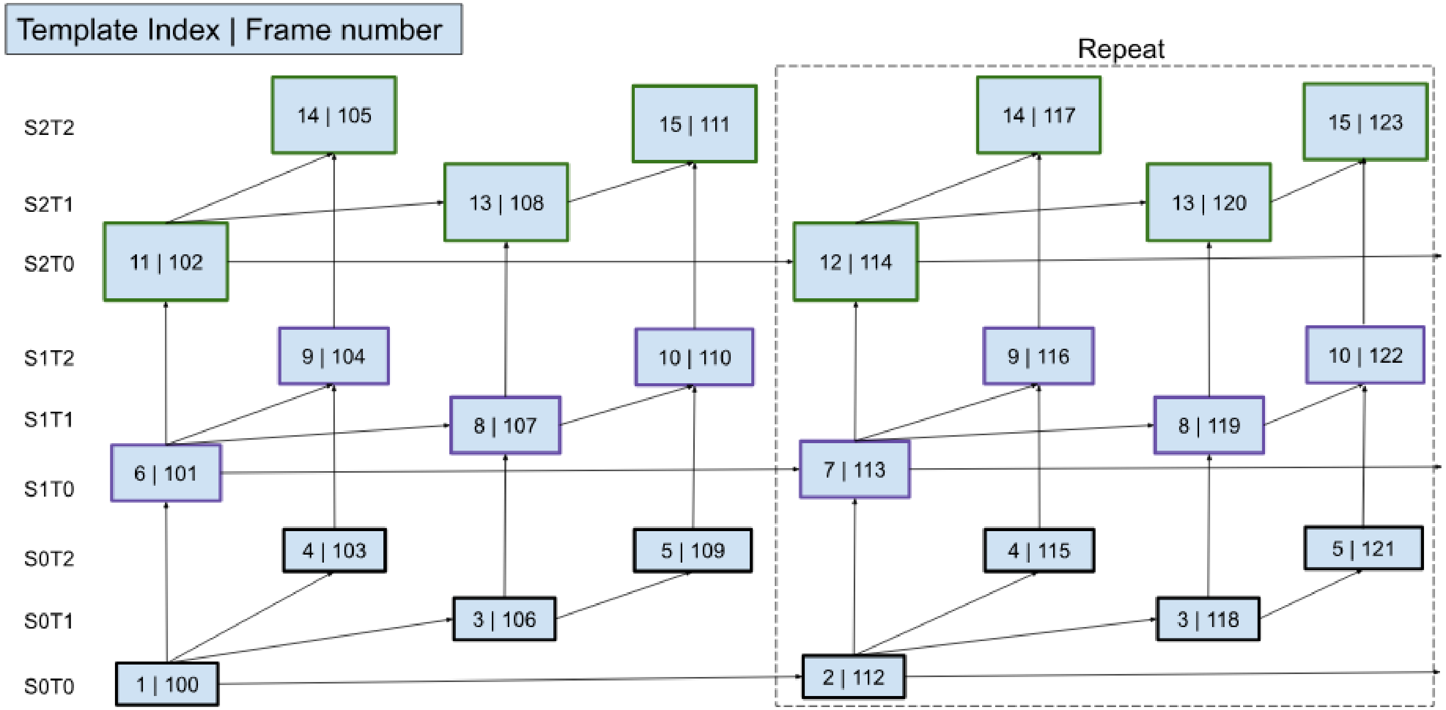
In this example, it is desired to send three simulcast encodings, each containing three temporal layers. When sending all encodings on a single SSRC, scalability mode 'S3T3' would be indicated within the scalability metadata OBU, and the Dependency Descriptor describes the dependency structure of all encodings.

When sending each simulcast encoding on a distinct SSRC, the scalability mode 'L1T3' would be indicated within the scalability metadata OBU of each bitstream, and the Dependency Descriptor in each stream describes only the dependency structure for that individual encoding. A distinct spatial_id (e.g. 0, 1, 2) could be used for each stream (if a single AV1 encoder is used to produce the three simulcast encodings), but if distinct AV1 encoders are used, the spatial_id values may not be distinct.

	Indication	Description	SFM behavior
DT0	Not present	F5 is not associated with DT0.	Do not forward F5 to a DT0 client.
DT1	Discardable	No frame in DT1 will reference F5.	Should forward F5 to a DT1 client, but may discard it, e.g., when bandwidth is low.
DT2	Switch	If it is possible to decode F5, then all future frames in DT2 are also decodable.	Forward F5 to the DT2 client. In addition, the SFM can rely on the fact that if the F5 frame is decodable for a particular client then that client may be switched to DT2.
DT3	Required	Future frames in DT3 reference F5.	Forward F5 to a DT3 client. No additional decisions can be made.

A.10.2.2 L3T3 Full SVC

This example uses three Chains. Chain 0 includes frames with spatial ID equal to 0 and temporal ID equal to 0. Chain 1 includes frames with spatial ID equal to 0 or 1 and temporal ID equal to 0. Chain 2 includes all frames with temporal ID equal to 0.



Idx	S	T	Fdiffs	Chains			Decode Targets								
				0	1	2	HD30 fps	HD15 fps	HD7.5 fps	VGA30 fps	VGA15 fps	VGA7.5fps	QVGA30 fps	QVGA15 fps	QVGA7.5 fps
1	0	0		0	0	0	S	S	S	S	S	S	S	S	S
2	0	0	12	12	11	10	R	R	R	R	R	R	S	S	S
3	0	1	6	6	5	4	R	R	-	R	R	-	S	D	-
4	0	2	3	3	2	1	R	-	-	R	-	-	D	-	-
5	0	2	3	9	8	7	R	-	-	R	-	-	D	-	-
6	1	0	1	1	1	1	S	S	S	S	S	S	-	-	-
7	1	0	12,1	1	1	1	R	R	R	S	S	S	-	-	-
8	1	1	6,1	7	6	5	R	R	-	S	D	-	-	-	-
9	1	2	3,1	4	3	2	R	-	-	D	-	-	-	-	-
10	1	2	3,1	10	9	8	R	-	-	D	-	-	-	-	-
11	2	0	1	2	1	1	S	S	S	-	-	-	-	-	-
12	2	0	12,1	2	1	1	S	S	S	-	-	-	-	-	-
13	2	1	6,1	8	7	6	S	D	-	-	-	-	-	-	-
14	2	2	3,1	5	4	3	D	-	-	-	-	-	-	-	-
15	2	2	3,1	11	10	9	D	-	-	-	-	-	-	-	-
decode_target_protected_by							2	2	2	1	1	1	0	0	0

(https://aomediacodec.github.io/av1-rtp-spec/.)

	Base layer The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **Enhancement layer** A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)* 6.2.3. OBU extension header semantics **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)*

	Note: The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). Coded video data of a temporal level with temporal_id T and spatial level with spatial_id S are only allowed to reference previously coded video data of temporal_id T' and spatial_id S', where T' <= T and S' <= S. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* If a coded video sequence contains at least one enhancement layer (OBUs with spatial_id greater than 0 or temporal_id greater than 0) then all frame headers and tile group OBUs associated with base (spatial_id equals 0 and temporal_id equals 0) and enhancement layer (spatial_id greater than 0 or temporal_id greater than 0) data must include the OBU extension header. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 205 of 669)*

Table 4. Webex Meetings Bandwidth per Resolution Table

Layer	Bandwidth Range
90p active thumbnail (each)	~60-100 kb/s
180p main video	125-200 kb/s
360p main video	470-640 kb/s
720p main video	900k-1.5 mb/s
Content sharing (sharpness, 1080p/5)	120k – 1.3 mb/s
Content sharing (motion, 720p/30)	900k – 2.5 mb/s

Webex Meetings Desktop App Bandwidth Controls

Webex administrators have 2 key controls to help control bandwidth as used by clients that connect to Webex meetings should they choose to. Namely, you can cap the meeting layouts at either 360p as the max available resolution, or to enable 720p layers. Whether your site is administered on Webex Control Hub or Webex Site Administrator, the following controls are available in Configuration > Common Site Settings > Options:

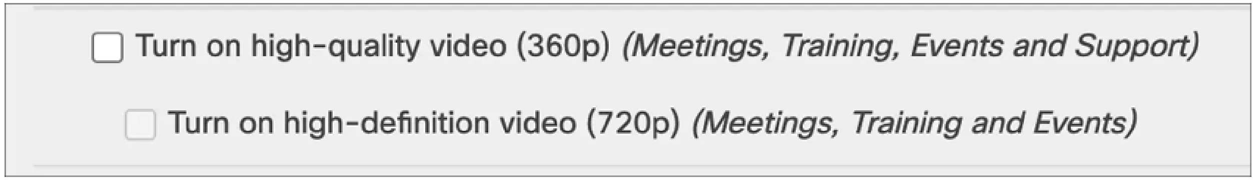


Figure 5.

Webex Meetings Desktop App Bandwidth Controls

Webex Media Improvements

The following are media improvements that have occurred in releases from 40.7 – 40.10

Improved Video Experience over WIFI (Defer video down-speeding): WIFI environments present a specific set of challenges and through testing we noticed the algorithms were too quick to drop resolution/bandwidth in certain WIFI environments. In 40.7 the algorithms have been updated to 'defer the down-speeding" of video for brief WIFI specific network dropouts. This resulted in improved user experience over WIFI. See figure5 for an illustration of this. When latency spikes for a specific duration the improved algorithms can detect that this is a WIFI specific event and not thus delay the down-speeding of the video resolution until that duration (n) is exceeded in which case would infer a different network event increasing latency.

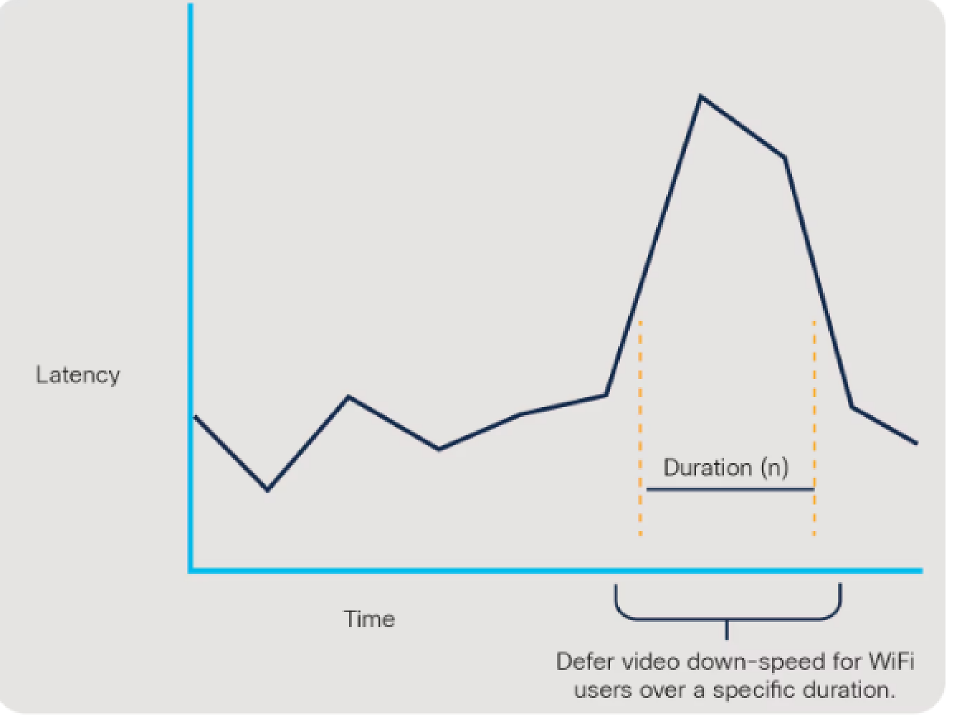


Figure 6.

Deferred Video Down-speeding

	Video Super Scaling is a technique used for users on sub-optimal networks. We provide 720p like experiences to users who are experiencing network impairments resulting in 360p connections to our servers for their meeting. Our updated scaling algorithm provides enhanced video quality when this degradation occurs, making network impairment downgrades less quality impacting. If a 360p image is simply scaled geometrically to a 720p resolution, the quality has not improved, but our technique involves migrating video post-processing tasks to the client for some video flows versus doing it in the cloud, resulting in being able to keep more precious bits of information to reconstruct the video at a given bandwidth. In a meeting as shown in figure 8 and 9, the bandwidth requirements are as follows. Assuming the receiver is not bandwidth restricted, Webex app will adapt to available network conditions. Table 8 outlines the bandwidth utilization. Figure 13 illustrates the Active Speaker layout which is the default meeting layout for Webex Teams App. **Table 8. Cisco Webex Teams Bandwidth Utilization** (table below) (https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings/white_paper_c11-691351.html.)

Table 8. Cisco Webex Teams Bandwidth Utilization

Meeting Scenario	Main Video+Audio (no content sharing)	Main Video+Audio (content sharing)
Webex Teams 1:1 Call	1.5 mb/s peak, 1.3 mb/s avg	1.5 mb/s peak, 1.3 mb/s avg
Webex Teams Meeting (fig 13 layout), 6 participants	2.3 mb/s peak, 1.7 mb/s avg	2.9/ mb/s peak, 2.0 mb/s avg
Webex Teams Meeting (fig 14 layout), 6 participants	2.5 mb/s peak, 1.8 mb/s avg	3.5 mb/s peak, 2.3 mb/s avg

wherein the scalable video coding router selectively forwards one or more of the set of enhancement layers, but fewer than all of the set of enhancement layers, to at least two of the remaining video receivers in the set of video receivers based upon the identified bandwidth-limited conditions;	The Accused Instrumentalities selectively forward one or more of the set of enhancement layers, but fewer than all of the set of enhancement layers, to at least two of the remaining video receivers in the set of video receivers based upon the identified bandwidth-limited conditions. For example, the AV1 standard discloses selectively forwarding (e.g., selective forwarding using SFU – selective forwarding unit or SFM – selective forwarding middlebox) one or more of the set of enhancement layers (e.g., enhancement layers), but fewer than all of the set of enhancement layers (e.g., enhancement layers), to at least two of the remaining video receivers in the set of video receivers (e.g., video data receiver such as a video bitstream decoder, etc.) based upon the identified bandwidth-limited conditions (e.g., network condition, available bandwidth condition for a receiving device, etc.). The AV1 standard is a video coding & decoding standard. It discloses that the AV1 standard is developed for video bitstream communication application like Internet related application, streaming, etc. It induces transmission of encoded video bitstream via Internet. It also discloses scaling according to varying bandwidth condition. Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) For example, one of the applications of the AV1 standard discloses a selective forwarding unit or middlebox (SFU or SFM) which analyses bandwidth condition between a publisher/transmitter (e.g., video router) and a playback

device (e.g., video receiver). According to the bandwidth condition, it selectively forwards layers to the playback device.

As shown below, it forwards only base layer to devices which are having low bandwidth available and other devices depending on their bandwidth conditions could receive different number of enhancement layers.

How Does SVC Work?

Video Router

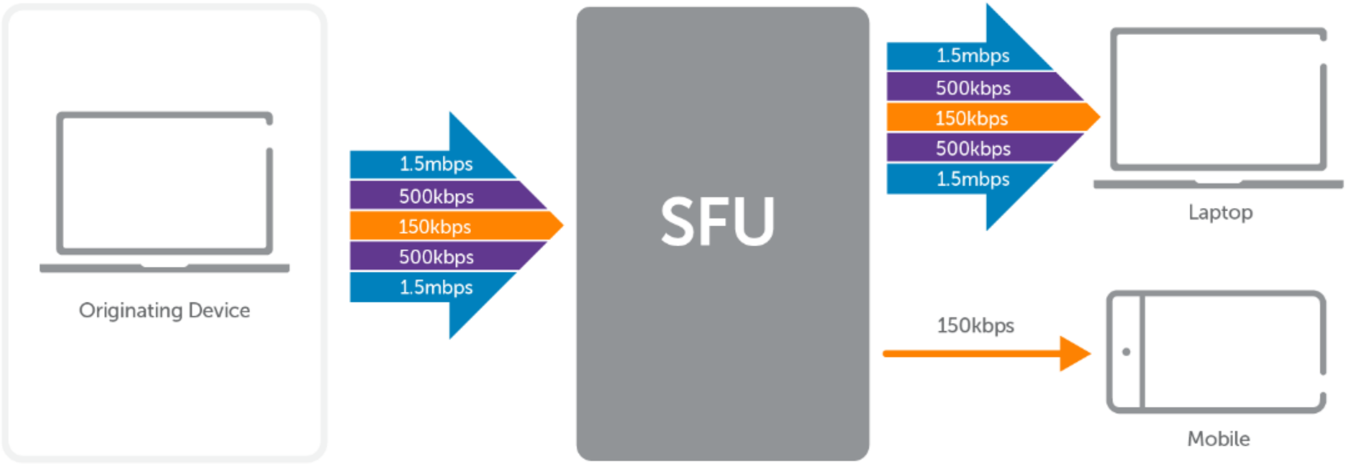
1. **Publisher**: Transmits a single media stream (with multiple video layers) to the Selective Forwarding Unit (SFU).

2. **SFU**: Examines the stream and selectively encodes it into high and low bitrate layers. It then distributes these layers based on each playback device's network limitations and available bandwidth.

3. **Playback Devices**: Receive the appropriate layers, ensuring optimal quality for each participant.

Video Receiver

(https://www.linkedin.com/pulse/scalable-video-coding-svc-sudeep-kumar-nofdc.)

	The SFU then sends the adjusted stream along to the appropriate **playback device**. Depending on the available bandwidth and other limitations of the target devices, some could receive a lower quality stream while others get the whole onion (so to speak).  (https://www.wowza.com/blog/scalable-video-coding-for-webrtc.)

	Selective Forwarding Middlebox (SFM) A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant. **Dependency Descriptor RTP Header Extension** **A.1 Introduction** This appendix describes the Dependency Descriptor (DD) RTP Header extension. The DD is used for conveying dependency information about individual video frames in a scalable video stream. The DD includes provisions for both temporal and spatial scalability. In the DD, the smallest unit for which dependencies are described is an RTP frame. An RTP frame contains one complete coded video frame and may also contain additional information (e.g., metadata). Each RTP frame is identified by a frame_number. When spatial scalability is used, there may be multiple RTP frames produced for the same time instant. Further, this specification allows for the transmission of an RTP frame over multiple packets. RTP frame aggregation is explicitly disallowed. Hereinafter, RTP frame will be referred to as frame. The DD uses the concept of Decode targets, each of which represents a subset of a scalable video stream necessary to decode the stream at a certain temporal and spatial fidelity. A frame may be associated with several Decode targets. This concept is used to facilitate selective forwarding, as is done by a Selective Forwarding Middlebox (SFM). Typically an SFM would select one Decode target for each endpoint, and forward all frames associated with that target.

6 MANE and SFM Behavior

If a packet contains an OBU with an OBU extension header then the entire packet is considered associated with the layer identified by the temporal_id and spatial_id combination that are indicated in the extension header. If a packet does not contain any OBU with an OBU extension header, then it is considered to be associated with all operating points.

The general function of a MANE or SFM is to selectively forward packets to receivers. To make forwarding decisions a MANE may inspect the media payload, so that it may need to be able to parse the AV1 bitstream and if so, cannot function when end-to-end encryption is enabled. An SFM does not parse the AV1 bitstream and therefore needs to obtain the information relevant to selective forwarding by other means, such as the Dependency Descriptor described in Appendix A. In addition to enabling bitstream-independent forwarding and support for end-to-end encryption, the Dependency Descriptor also enables forwarding where the metadata OBU provided in the AV1 bitstream is not sufficient to express the structure of the stream.

6.1.1 Example

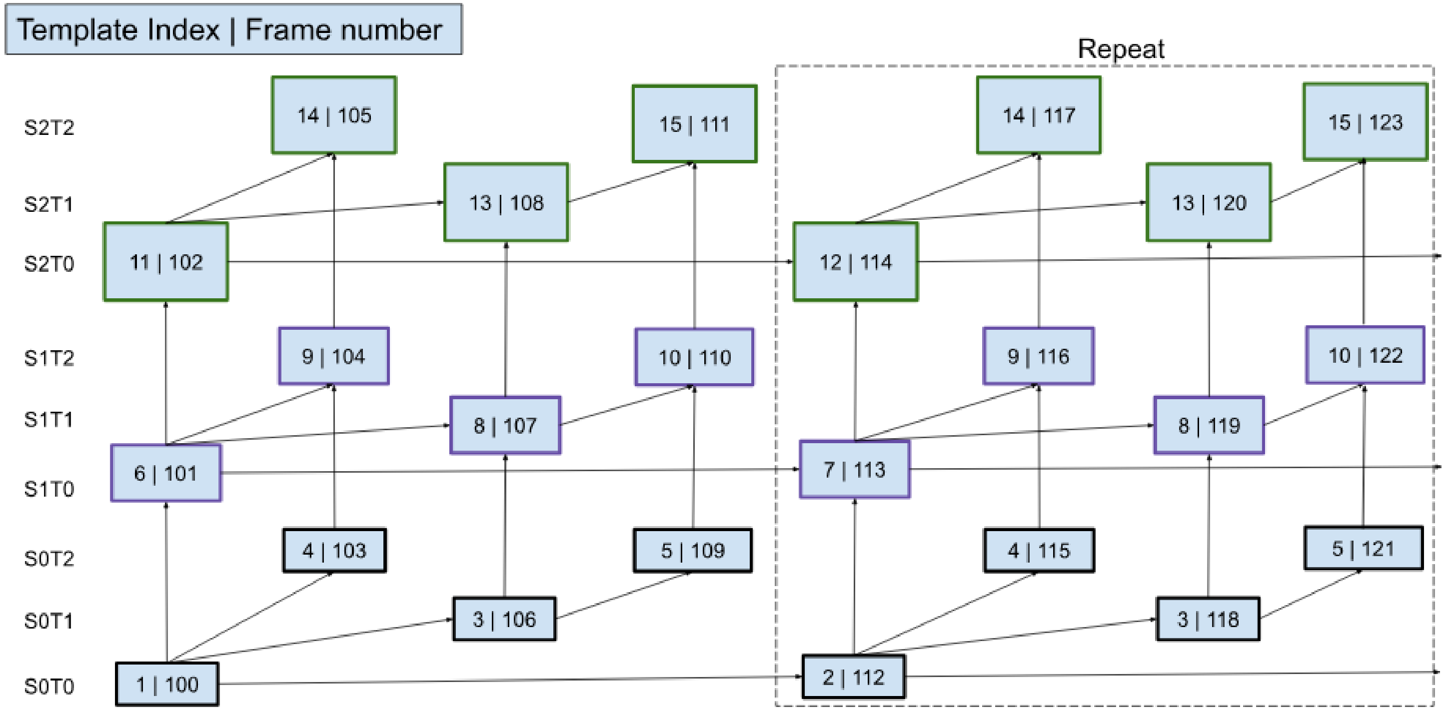
In this example, it is desired to send three simulcast encodings, each containing three temporal layers. When sending all encodings on a single SSRC, scalability mode 'S3T3' would be indicated within the scalability metadata OBU, and the Dependency Descriptor describes the dependency structure of all encodings.

When sending each simulcast encoding on a distinct SSRC, the scalability mode 'L1T3' would be indicated within the scalability metadata OBU of each bitstream, and the Dependency Descriptor in each stream describes only the dependency structure for that individual encoding. A distinct spatial_id (e.g. 0, 1, 2) could be used for each stream (if a single AV1 encoder is used to produce the three simulcast encodings), but if distinct AV1 encoders are used, the spatial_id values may not be distinct.

	Indication	Description	SFM behavior
DT0	Not present	F5 is not associated with DT0.	Do not forward F5 to a DT0 client.
DT1	Discardable	No frame in DT1 will reference F5.	Should forward F5 to a DT1 client, but may discard it, e.g., when bandwidth is low.
DT2	Switch	If it is possible to decode F5, then all future frames in DT2 are also decodable.	Forward F5 to the DT2 client. In addition, the SFM can rely on the fact that if the F5 frame is decodable for a particular client then that client may be switched to DT2.
DT3	Required	Future frames in DT3 reference F5.	Forward F5 to a DT3 client. No additional decisions can be made.

A.10.2.2 L3T3 Full SVC

This example uses three Chains. Chain 0 includes frames with spatial ID equal to 0 and temporal ID equal to 0. Chain 1 includes frames with spatial ID equal to 0 or 1 and temporal ID equal to 0. Chain 2 includes all frames with temporal ID equal to 0.



Idx	S	T	Fdiffs	Chains			Decode Targets								
				0	1	2	HD30 fps	HD15 fps	HD7.5 fps	VGA30 fps	VGA15 fps	VGA7.5fps	QVGA30 fps	QVGA15 fps	QVGA7.5 fps
1	0	0		0	0	0	S	S	S	S	S	S	S	S	S
2	0	0	12	12	11	10	R	R	R	R	R	R	S	S	S
3	0	1	6	6	5	4	R	R	-	R	R	-	S	D	-
4	0	2	3	3	2	1	R	-	-	R	-	-	D	-	-
5	0	2	3	9	8	7	R	-	-	R	-	-	D	-	-
6	1	0	1	1	1	1	S	S	S	S	S	S	-	-	-
7	1	0	12,1	1	1	1	R	R	R	S	S	S	-	-	-
8	1	1	6,1	7	6	5	R	R	-	S	D	-	-	-	-
9	1	2	3,1	4	3	2	R	-	-	D	-	-	-	-	-
10	1	2	3,1	10	9	8	R	-	-	D	-	-	-	-	-
11	2	0	1	2	1	1	S	S	S	-	-	-	-	-	-
12	2	0	12,1	2	1	1	S	S	S	-	-	-	-	-	-
13	2	1	6,1	8	7	6	S	D	-	-	-	-	-	-	-
14	2	2	3,1	5	4	3	D	-	-	-	-	-	-	-	-
15	2	2	3,1	11	10	9	D	-	-	-	-	-	-	-	-
decode_target_protected_by							2	2	2	1	1	1	0	0	0

(https://aomediacodec.github.io/av1-rtp-spec/.)

	Base layer The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **Enhancement layer** A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)* 6.2.3. OBU extension header semantics **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)*

	Note: The term "spatial" refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). Coded video data of a temporal level with temporal_id T and spatial level with spatial_id S are only allowed to reference previously coded video data of temporal_id T' and spatial_id S', where T' <= T and S' <= S. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* If a coded video sequence contains at least one enhancement layer (OBUs with spatial_id greater than 0 or temporal_id greater than 0) then all frame headers and tile group OBUs associated with base (spatial_id equals 0 and temporal_id equals 0) and enhancement layer (spatial_id greater than 0 or temporal_id greater than 0) data must include the OBU extension header. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 205 of 669)*

Table 4. Webex Meetings Bandwidth per Resolution Table

Layer	Bandwidth Range
90p active thumbnail (each)	~60-100 kb/s
180p main video	125-200 kb/s
360p main video	470-640 kb/s
720p main video	900k-1.5 mb/s
Content sharing (sharpness, 1080p/5)	120k – 1.3 mb/s
Content sharing (motion, 720p/30)	900k – 2.5 mb/s

Webex Meetings Desktop App Bandwidth Controls

Webex administrators have 2 key controls to help control bandwidth as used by clients that connect to Webex meetings should they choose to. Namely, you can cap the meeting layouts at either 360p as the max available resolution, or to enable 720p layers. Whether your site is administered on Webex Control Hub or Webex Site Administrator, the following controls are available in Configuration > Common Site Settings > Options:

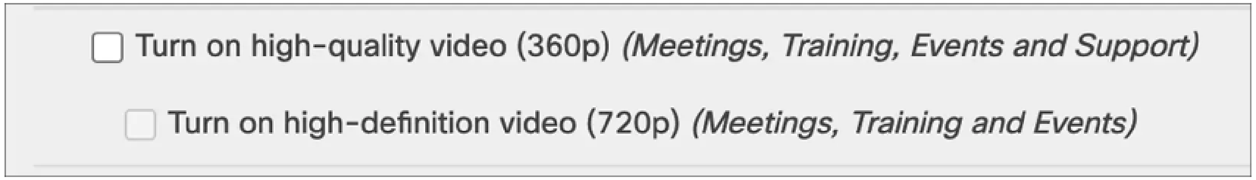


Figure 5.

Webex Meetings Desktop App Bandwidth Controls

Webex Media Improvements

The following are media improvements that have occurred in releases from 40.7 – 40.10

Improved Video Experience over WIFI (Defer video down-speeding): WIFI environments present a specific set of challenges and through testing we noticed the algorithms were too quick to drop resolution/bandwidth in certain WIFI environments. In 40.7 the algorithms have been updated to 'defer the down-speeding" of video for brief WIFI specific network dropouts. This resulted in improved user experience over WIFI. See figure5 for an illustration of this. When latency spikes for a specific duration the improved algorithms can detect that this is a WIFI specific event and not thus delay the down-speeding of the video resolution until that duration (n) is exceeded in which case would infer a different network event increasing latency.

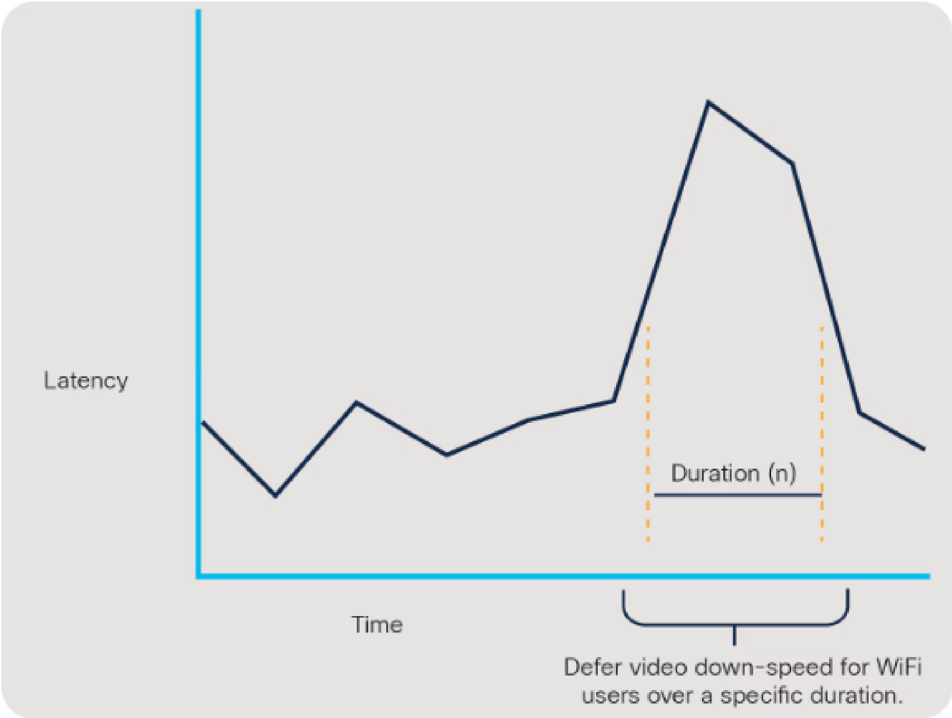


Figure 6.

Deferred Video Down-speeding

	Video Super Scaling is a technique used for users on sub-optimal networks. We provide 720p like experiences to users who are experiencing network impairments resulting in 360p connections to our servers for their meeting. Our updated scaling algorithm provides enhanced video quality when this degradation occurs, making network impairment downgrades less quality impacting. If a 360p image is simply scaled geometrically to a 720p resolution, the quality has not improved, but our technique involves migrating video post-processing tasks to the client for some video flows versus doing it in the cloud, resulting in being able to keep more precious bits of information to reconstruct the video at a given bandwidth. In a meeting as shown in figure 8 and 9, the bandwidth requirements are as follows. Assuming the receiver is not bandwidth restricted, Webex app will adapt to available network conditions. Table 8 outlines the bandwidth utilization. Figure 13 illustrates the Active Speaker layout which is the default meeting layout for Webex Teams App. **Table 8. Cisco Webex Teams Bandwidth Utilization**

Meeting Scenario	Main Video+Audio (no content sharing)	Main Video+Audio (content sharing)
Webex Teams 1:1 Call	1.5 mb/s peak, 1.3 mb/s avg	1.5 mb/s peak, 1.3 mb/s avg
Webex Teams Meeting (fig 13 layout), 6 participants	2.3 mb/s peak, 1.7 mb/s avg	2.9/ mb/s peak, 2.0 mb/s avg
Webex Teams Meeting (fig 14 layout), 6 participants	2.5 mb/s peak, 1.8 mb/s avg	3.5 mb/s peak, 2.3 mb/s avg

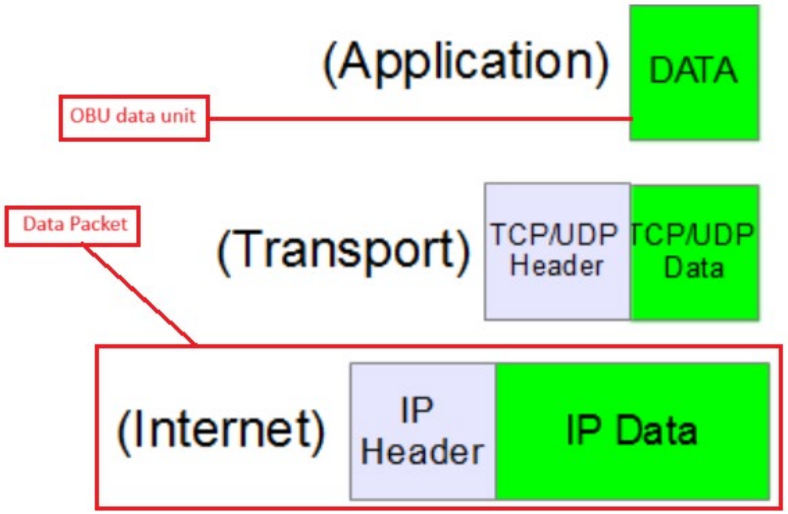
(https://www.cisco.com/c/en/us/products/collateral/conferencing/webex-meetings/white_paper_c11-691351.html.)

wherein the layered video data stream is transmitted according to an internet protocol,	The Accused Instrumentalities include a scalable video coding router, wherein the layered video data stream is transmitted according to an internet protocol. For example, the AV1 standard discloses the method such that the layered video data stream (e.g., scalable video bitstream, etc.) is transmitted according to an internet protocol (e.g., Internet, etc.). Work within AOMedia is organized in Working Groups, each with a defined scope and deliverables. For example, the Codec Working Group is tasked with developing video coding standards and manages the AV1 standard. AV1, which was designed from the get-go for video on the Web, was the initial project of AOMedia and was published in 2018. Work has since expanded to include immersive sound, starting with IAMF, and volumetric media for high-quality 3D graphics. (https://aomedia.org/about/story/.) **Selective Forwarding Middlebox (SFM)** A middlebox that relays streams among transmitting and receiving clients by selectively forwarding packets without accessing the media (RFC7667). **Temporal unit** Defined by the AV1 specification: A temporal unit consists of all the OBUs that are associated with a specific, distinct time instant. (https://aomediacodec.github.io/av1-rtp-spec/.)

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.)
and wherein each layer of the layered video data stream comprises data packets, each	The Accused Instrumentalities include a scalable video coding router, wherein each layer of the layered video data stream comprises data packets, each of which is encoded with a sequence number and a layer identifier, and wherein the layer identifier for each data packet is based upon a layer to which the packet belongs.

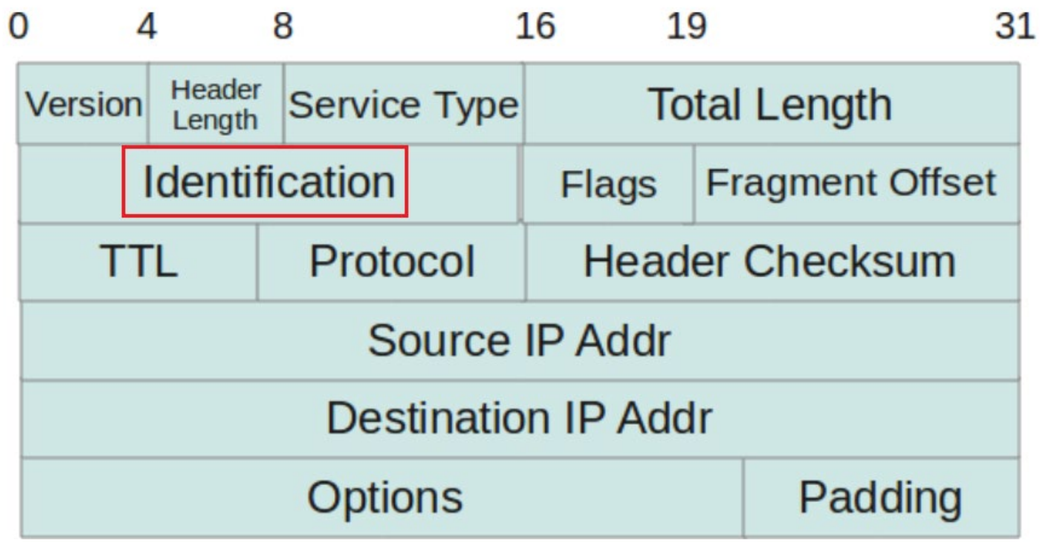
of which is encoded with a sequence number and a layer identifier, and wherein the layer identifier for each data packet is based upon a layer to which the packet belongs.	For example, the AV1 standard discloses the method such that each layer of the layered video data stream comprises data packets (e.g., IP packets of video bitstream data units) each of which is encoded with a sequence number (e.g., identification number value of IP packet, etc.) and a layer identifier (e.g., a layer identifier such as base layer, enhancement layer, etc.) and wherein the layer identifier (e.g., the layer identifier such as base layer, enhancement layer, etc.) for each data packet (e.g., IP packets of video bitstream data units) is based upon a layer (e.g., layer such as base layer, enhancement layer, etc.) to which the packet belongs. As shown below, the AV1 standard discloses an encoded video data bitstream using scalable video coding in a sequence of OBUs i.e., open bitstream unit. The OBU data units are transmitted in packetized format over Internet. These packets are governed by IP protocol and communicated as an IP packet. An IP packet comprises a header part and a payload/data part. The header part of the IP packet comprises an identification field which denotes a sequence number of IP packets i.e., data packets, transmitted.

	AV1 Features **ROYALTY-FREE** Interoperable and open **UBIQUITOUS** Scales to any modern device at any bandwith **FLEXIBLE** For use in both commercial and non-commercial content, including user-generated content **30% BETTER COMPRESSION *** Uses less data while delivering 4k UHD video and beyond when compared to alternatives **OPTIMIZED** Developed for the internet and related applications and services-from browsers and streaming to videoconferencing services (https://aomedia.org/av1-features/.) **Bitstream** The sequence of bits generated by encoding a sequence of frames. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)*

	OBU All structures are packetized in "Open Bitstream Units" or OBUs. Each OBU has a header, which provides identifying information for the contained data (payload). *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page – 4 of 669)* **Open Bitstream Unit (OBU)** The smallest bitstream data framing unit in AV1. All AV1 bitstream structures are packetized in OBUs. (https://aomediacodec.github.io/av1-rtp-spec/#3-media-format-description.)  (https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

	These are a set of standard rules that allows different types of computers to communicate with each other. The IP protocol ensures that each computer that is connected to the Internet is having a specific serial number called the IP address. TCP specifies how data is exchanged over the internet and how it should be broken into IP packets. It also makes sure that the packets have information about the source of the message data, the destination of the message data, the sequence in which the message data should be re-assembled, and checks if the message has been sent correctly to the specific destination. The TCP is also known as a connection-oriented protocol. (https://www.geeksforgeeks.org/types-of-internet-protocols/.)

IP Header



(https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

- ***Identification(16 bits)***: This field is used for uniquely identifying the IP datagrams. This value is incremented every-time an IP datagram is sent from source to the destination. This field comes in handy while reassembly of fragmented IP data grams.

(https://www.thegeekstuff.com/2012/03/ip-protocol-header/.)

Further, the OBU data unit comprises a metadata syntax which discloses scalability corresponding to the OBU. It discloses three types of scalabilities i.e., Spatial scalability, Temporal scalability and Quality scalability. These

	scalabilities define a spatial layer having a corresponding spatial_id and a temporal layer having a corresponding temporal_id. Further, the AV1 standard discloses deriving a layered coded bitstream of base layer and enhancement layers using scalable video coding. It discloses a base layer having both spatial_id and temporal_id equal to zero and enhancement layers with at least one of spatial_id or temporal_id values greater than zero. **Bitstream** The sequence of bits generated by encoding a sequence of frames. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **5.8.1. General metadata OBU syntax** (see table below) *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 34 of 669)*

5.8.1. General metadata OBU syntax

metadata_obu( ) {	Type
metadata_type	leb128()
if (metadata_type == METADATA_TYPE_ITUT_T35)	
metadata_itut_t35( )	
else if (metadata_type == METADATA_TYPE_HDR_CLL)	
metadata_hdr_cll( )	
else if (metadata_type == METADATA_TYPE_HDR_MDCV)	
metadata_hdr_mdcv( )	
else if (metadata_type == METADATA_TYPE_SCALABILITY)	
metadata_scalability( )	
else if (metadata_type == METADATA_TYPE_TIMECODE)	
metadata_timecode( )	
}	

	5.8.5. Metadata scalability syntax
	
	metadata_scalability( ) { — Type
	    scalability_mode_idc — f(8)
	    if (scalability_mode_idc == SCALABILITY_SS)
	        scalability_structure( )
	}
	
	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

5.8.6. Scalability structure syntax

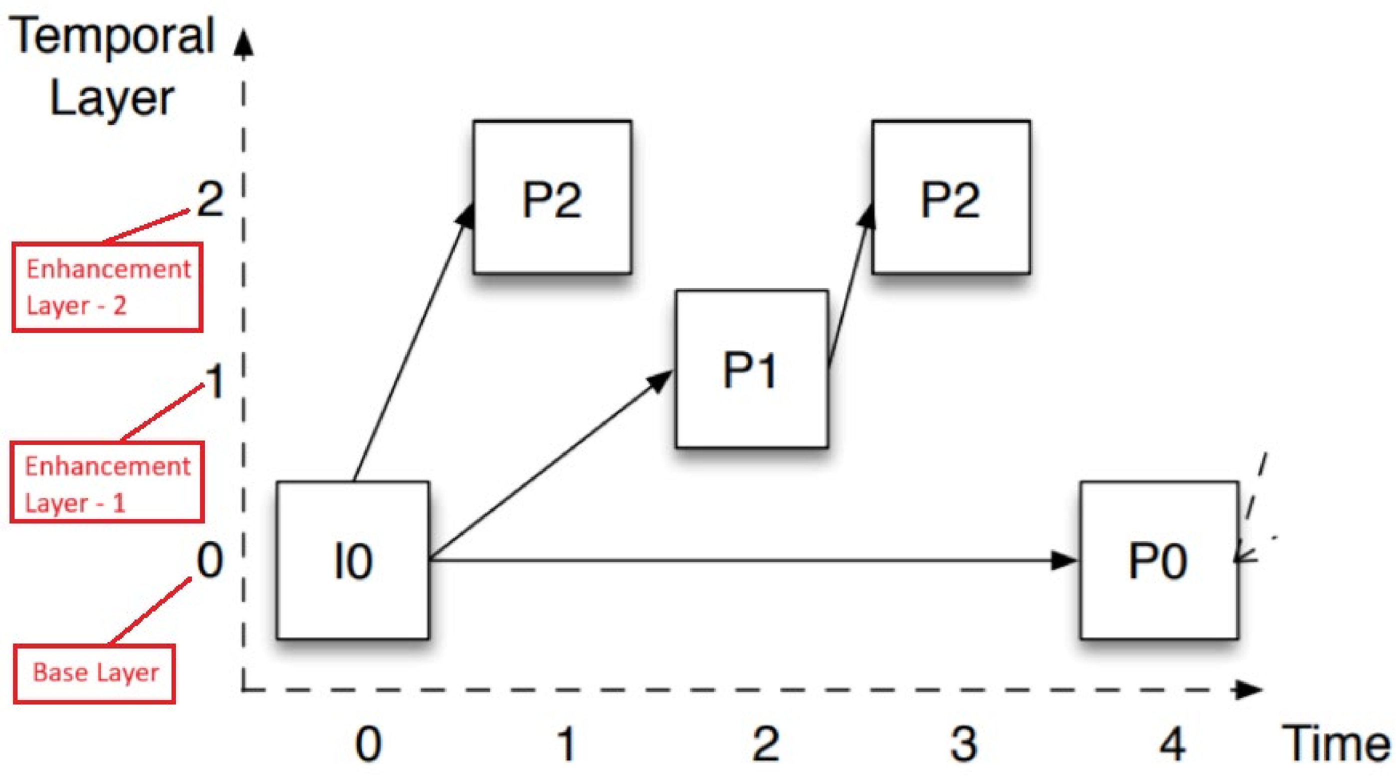
scalability_structure( ) {	Type
spatial_layers_cnt_minus_1	f(2)
spatial_layer_dimensions_present_flag	f(1)
spatial_layer_description_present_flag	f(1)
temporal_group_description_present_flag	f(1)
scalability_structure_reserved_3bits	f(3)
if (spatial_layer_dimensions_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1 ; i++) {	
spatial_layer_max_width[i]	f(16)
spatial_layer_max_height[i]	f(16)
}	
}	
if (spatial_layer_description_present_flag) {	
for (i = 0; i <= spatial_layers_cnt_minus_1; i++)	
spatial_layer_ref_id[i]	f(8)
}	
if (temporal_group_description_present_flag) {	
temporal_group_size	f(8)
for (i = 0; i < temporal_group_size; i++) {	
temporal_group_temporal_id[i]	f(3)
temporal_group_temporal_switching_up_point_flag[i]	f(1)
temporal_group_spatial_switching_up_point_flag[i]	f(1)
temporal_group_ref_cnt[i]	f(3)
for (j = 0; j < temporal_group_ref_cnt[i]; j++) {	
temporal_group_ref_pic_diff[i][j]	f(8)
}	

Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 36 of 669)

spatial_layer_ref_id[i] specifies the spatial_id value of the frame within the current temporal unit that the frame of layer i uses for reference. If no frame within the current temporal unit is used for reference the value must be equal to 255.

	Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669) Note that for a given picture, all frames follow the same inter-picture temporal dependency structure. However, the frame rate of each layer can be different from each other. The specified dependency structure in the scalability structure data must be for the highest frame rate layer. **temporal_group_temporal_id[i]** specifies the temporal_id value for the i-th picture in the temporal group. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 137 of 669)* **temporal_id** specifies the temporal level of the data contained in the OBU. When temporal_id is not present, temporal_id is inferred to be equal to 0. **spatial_id** specifies the spatial level of the data contained in the OBU. When spatial_id is not present, spatial_id is inferred to be equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 111 of 669)* **Base layer** The layer with spatial_id and temporal_id values equal to 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 1 of 669)* **Enhancement layer** A layer with either spatial_id greater than 0 or temporal_id greater than 0. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 2 of 669)*

	Note: The term “spatial” refers to the fact that the enhancement here occurs in the spatial dimension: either as an increase in spatial resolution, or an increase in spatial fidelity (increased SNR). Tile group OBU data associated with spatial_id and temporal_id equal to 0 are referred to as the base layer, whereas tile group OBU data that are associated with spatial_id greater than 0 or temporal_id greater than 0 are referred to as enhancement layer(s). Coded video data of a temporal level with temporal_id T and spatial level with spatial_id S are only allowed to reference previously coded video data of temporal_id T’ and spatial_id S’, where T’ <= T and S’ <= S. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 112 of 669)* **Note:** Examples are given for non-scalable cases, but the constraints also apply to each operating point of a scalable stream. For example, consider a 60fps spatial scalable stream with a base layer at 960x540, and an enhancement layer at 1920x1080. The operating point containing just the base layer may be labelled as level 3.0, while the operating point containing both the base and enhancement layer may be labelled as level 4.1. *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 641 of 669)*

	 *Source: AV1 Bitstream & Decoding Process Specification Version 1.0.0 with Errata 1 – 2018 (Page 126 of 669)*

Dependency Descriptor RTP Header Extension

A.1 Introduction

This appendix describes the Dependency Descriptor (DD) RTP Header extension. The DD is used for conveying dependency information about individual video frames in a scalable video stream. The DD includes provisions for both temporal and spatial scalability.

In the DD, the smallest unit for which dependencies are described is an RTP frame. An RTP frame contains one complete coded video frame and may also contain additional information (e.g., metadata). Each RTP frame is identified by a frame_number. When spatial scalability is used, there may be multiple RTP frames produced for the same time instant. Further, this specification allows for the transmission of an RTP frame over multiple packets. RTP frame aggregation is explicitly disallowed. Hereinafter, RTP frame will be referred to as frame.

The DD uses the concept of Decode targets, each of which represents a subset of a scalable video stream necessary to decode the stream at a certain temporal and spatial fidelity. A frame may be associated with several Decode targets. This concept is used to facilitate selective forwarding, as is done by a Selective Forwarding Middlebox (SFM). Typically an SFM would select one Decode target for each endpoint, and forward all frames associated with that target.

A.8 Dependency Descriptor Format

To facilitate the work of selectively forwarding portions of a scalable video bitstream, as is done by an SFM, for each packet, the following information is made available (even though not all elements are present in every packet).

- spatial ID
- temporal ID
- DTIs
- frame_number of the current frame
- frame_number of each of the Referred frames
- frame_number of last frame in each Chain

6 MANE and SFM Behavior

If a packet contains an OBU with an OBU extension header then the entire packet is considered associated with the layer identified by the temporal_id and spatial_id combination that are indicated in the extension header. If a packet does not contain any OBU with an OBU extension header, then it is considered to be associated with all operating points.

The general function of a MANE or SFM is to selectively forward packets to receivers. To make forwarding decisions a MANE may inspect the media payload, so that it may need to be able to parse the AV1 bitstream and if so, cannot function when end-to-end encryption is enabled. An SFM does not parse the AV1 bitstream and therefore needs to obtain the information relevant to selective forwarding by other means, such as the Dependency Descriptor described in Appendix A. In addition to enabling bitstream-independent forwarding and support for end-to-end encryption, the Dependency Descriptor also enables forwarding where the metadata OBU provided in the AV1 bitstream is not sufficient to express the structure of the stream.

6.1.1 Example

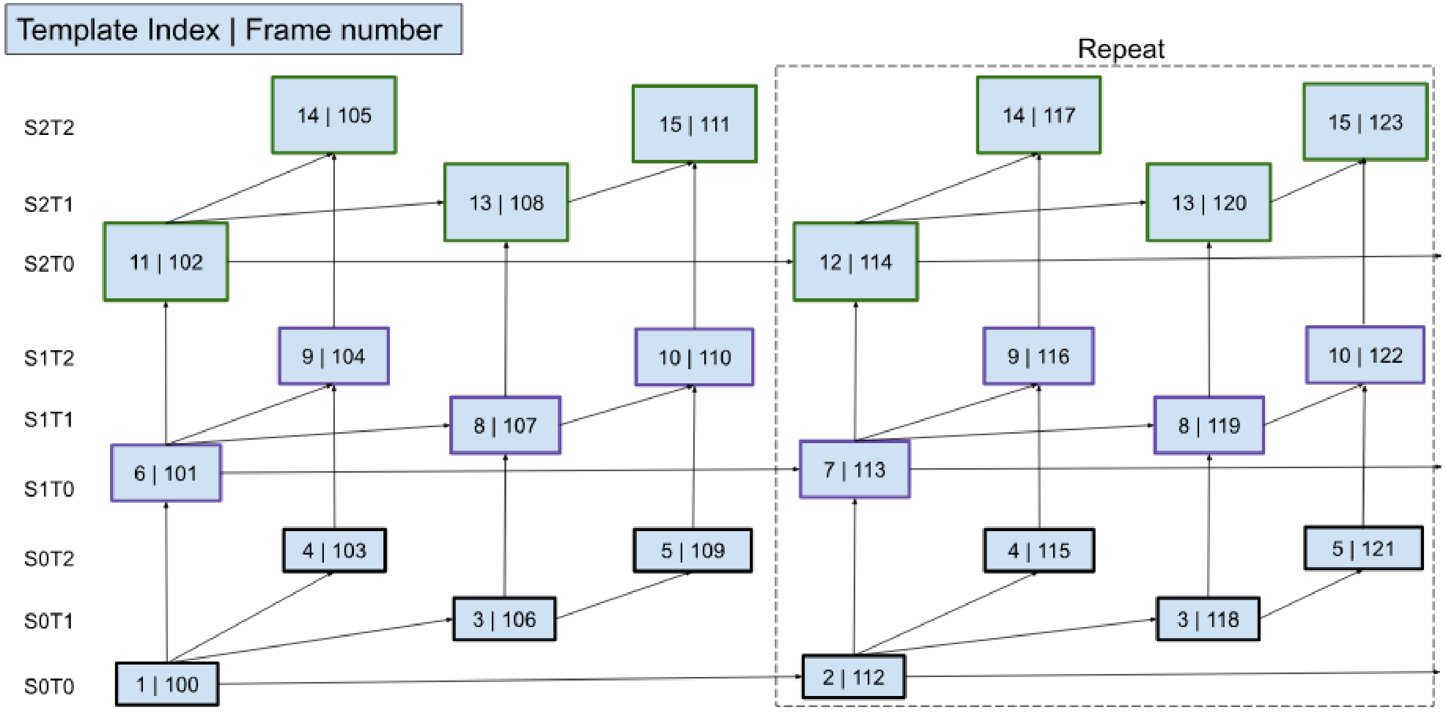
In this example, it is desired to send three simulcast encodings, each containing three temporal layers. When sending all encodings on a single SSRC, scalability mode 'S3T3' would be indicated within the scalability metadata OBU, and the Dependency Descriptor describes the dependency structure of all encodings.

When sending each simulcast encoding on a distinct SSRC, the scalability mode 'L1T3' would be indicated within the scalability metadata OBU of each bitstream, and the Dependency Descriptor in each stream describes only the dependency structure for that individual encoding. A distinct spatial_id (e.g. 0, 1, 2) could be used for each stream (if a single AV1 encoder is used to produce the three simulcast encodings), but if distinct AV1 encoders are used, the spatial_id values may not be distinct.

	Indication	Description	SFM behavior
DT0	Not present	F5 is not associated with DT0.	Do not forward F5 to a DT0 client.
DT1	Discardable	No frame in DT1 will reference F5.	Should forward F5 to a DT1 client, but may discard it, e.g., when bandwidth is low.
DT2	Switch	If it is possible to decode F5, then all future frames in DT2 are also decodable.	Forward F5 to the DT2 client. In addition, the SFM can rely on the fact that if the F5 frame is decodable for a particular client then that client may be switched to DT2.
DT3	Required	Future frames in DT3 reference F5.	Forward F5 to a DT3 client. No additional decisions can be made.

A.10.2.2 L3T3 Full SVC

This example uses three Chains. Chain 0 includes frames with spatial ID equal to 0 and temporal ID equal to 0. Chain 1 includes frames with spatial ID equal to 0 or 1 and temporal ID equal to 0. Chain 2 includes all frames with temporal ID equal to 0.



Idx	S	T	Fdiffs	Chains			Decode Targets								
				0	1	2	HD30 fps	HD15 fps	HD7.5 fps	VGA30 fps	VGA15 fps	VGA7.5fps	QVGA30 fps	QVGA15 fps	QVGA7.5 fps
1	0	0		0	0	0	S	S	S	S	S	S	S	S	S
2	0	0	12	12	11	10	R	R	R	R	R	R	S	S	S
3	0	1	6	6	5	4	R	R	-	R	R	-	S	D	-
4	0	2	3	3	2	1	R	-	-	R	-	-	D	-	-
5	0	2	3	9	8	7	R	-	-	R	-	-	D	-	-
6	1	0	1	1	1	1	S	S	S	S	S	S	-	-	-
7	1	0	12,1	1	1	1	R	R	R	S	S	S	-	-	-
8	1	1	6,1	7	6	5	R	R	-	S	D	-	-	-	-
9	1	2	3,1	4	3	2	R	-	-	D	-	-	-	-	-
10	1	2	3,1	10	9	8	R	-	-	D	-	-	-	-	-
11	2	0	1	2	1	1	S	S	S	-	-	-	-	-	-
12	2	0	12,1	2	1	1	S	S	S	-	-	-	-	-	-
13	2	1	6,1	8	7	6	S	D	-	-	-	-	-	-	-
14	2	2	3,1	5	4	3	D	-	-	-	-	-	-	-	-
15	2	2	3,1	11	10	9	D	-	-	-	-	-	-	-	-
decode_target_protected_by							2	2	2	1	1	1	0	0	0

(https://aomediacodec.github.io/av1-rtp-spec/.)